Safeguarding data protection in an open data world
Dalla Corte, Lorenzo

Document version:
Publisher's PDF, also known as Version of record

Publication date:
2020

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Safeguarding Data Protection in an Open Data World:
On the idea of balancing open data and data protection in the development of the smart city environment

Proefschrift ter verkrijging van de graad van doctor aan Tilburg University, op gezag van de rector magnificus, prof. dr. K. Sijtsma, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie aan Tilburg University op maandag 18 mei 2020 om 13.30 uur

door
Lorenzo Dalla Corte,

tegenwoordig in Feltre, Italië
Promotor: prof. dr. E. Kosta
Copromotor: dr. ir. B. van Loenen
Leden promotiecommissie: prof. dr. T. Scassa
                      prof. A. Mantelerò
                      prof. dr. R.E. Leenes
                      dr. L.E.M. Taylor
Acknowledgements

I am bound to acknowledge, first and foremost, my parents, Maura and Franco. Writing this dissertation would not have been possible without them and their unwavering care. Good job, mom and dad!

I must then thank my supervisor, dr. Bastiaan van Loenen, and my promotor, prof. dr. Eleni Kosta: without their infinite patience none of this would have been achievable. I cannot convey how lucky I was at having had the chance of working with them. Thank you, Eleni, and thank you, Bastiaan: I could not have wished for better supervisors.

I would be remiss if I did not recognise TILT’s director, prof. dr. Ronald Leenes, one of the most influential people in my professional development. Thank you, Ronald, I don’t think I would have completed this dissertation if it were not for you and your (often involuntary or anecdote-based) life lessons.

Speaking of TILT and of (current and former) TILTies: thank you guys, the whole lot of you. I would like to acknowledge, in particular, Silvia De Conca, Irene Kamara, and Mara Paun, whose friendship kept me tethered to sanity, and who never got tired to talk about cuisine with me: just say the word and we’re opening that bistrot. I would also like to thank the other two components of Tiny Office: Karine e Silva, who is likely to be the best dog mom in the world, and Jingze Li, who once made me eat jellyfish. My gratitude also goes to Claudia Quelle, Shaz Jameson, Aviva De Groot, and Sascha van Schendel, who read great books and watch good movies, and to Dimitra Stefanatou, who taught me the ropes of this job.

I would also like to thank TU Delft’s Kennniscentrum Open Data, whose diverse expertise was fundamental for my research, and all its members, particularly dr. Frederika Welle Donker, prof. Hendrik Ploeger, Agung Indrajit, and Warakan Supinajaroen.

Besides my two academic homes, Tilburg University and TU Delft, I had a third: the Computers, Privacy, and Data Protection (CPDP) conferences. I would like to thank them too, particularly Paul De Hert, Rosamunde Van Brakel, Dara Hallinan, and Imge Ozcan. CPDP has been fundamental in developing my understanding of (and love for) privacy and data protection, and having had the chance of giving back a bit made me happier than you can possibly know.

Several others have been extremely important, either personally, professionally, or both. They are too many to list, but I would like to mention Damian Clifford, and our nightly chats about the essence of the right to data protection, which I hope will never be made object of a subpoena, and Jef Ausloos, with whom I raided several Szechuanese restaurants while discussing the notion of proportionality. My thanks also go to Federico Budel, who has been calling be a monster for the past 25 years.

I must also acknowledge the numerous scholars on whose shoulders I was allowed to stand while writing this thesis. I cannot name them all, as the fields of privacy and data protection are burgeoning, I already acknowledged some of them, and most can be found in this dissertation’s references anyway. Amongst many, I would like to thank Alessandro Mantelero, Bert-Jaap Koops, Frederik Borgesius, Gloria González Fuster, Lee Bygrave, Lilian Edwards, Linnet Taylor, Michael Veale, Nadya Purtova, Nóra Ni Loideain, Orla Lynskey, Raphael Gellert, Serge Gutwirth, and Teresa Scassa. There are surely more, but I am running out of space, and memory has never been my strongest suit.

Finally, this research was performed with the financial support of the Dutch STW/NWO Maps4Society program (project number 13718), which I thank.
# TABLE OF CONTENTS

1. INTRODUCTION

1.1 Smart cities, open data, and data protection ............................................. 7
1.2 Research questions and significance ......................................................... 13
1.3 Methodology .............................................................................................. 16
1.4 Outline ....................................................................................................... 20

2. ON THE CONCEPT OF ‘SMART CITY’ ......................................................... 22

2.1 Making sense of the smart city .................................................................. 23
2.2 What is a smart city? .................................................................................. 25
  2.2.1 Perspectives on the smart city ............................................................ 26
  2.2.1.1 Technological perspective on smart cities ..................................... 26
  2.2.1.2 Organisational perspective on smart cities: governance and management 27
  2.2.1.3 Human perspective on smart cities: human capital ..................... 28
  2.2.2 The smart city as a regulatory instrument ....................................... 30
  2.2.2.1 Top-down regulatory capacity ....................................................... 31
  2.2.2.2 Bottom-up regulatory capacity ....................................................... 33
  2.2.3 Visions of a smarter city ..................................................................... 35
  2.2.4 Implementation of the smart city ......................................................... 37
  2.2.5 Teleology of the smart city .................................................................. 40
  2.2.6 A jungle of standards .......................................................................... 41
  2.2.7 Defining smart cities? ......................................................................... 43
2.3 Smart cities, privacy, and data protection ................................................. 46
  2.3.1 Datafication and data protection ......................................................... 47
  2.3.2 Values and design ................................................................................ 49
  2.3.2.1 Which values? ................................................................................ 50
  2.3.2.2 But how? Implementing privacy and data protection in design ........ 57
2.4 Smart cities and open data ........................................................................ 61
2.5 Conclusion .................................................................................................. 64

3. OPEN DATA .................................................................................................. 67

3.1 What is open data? ..................................................................................... 67
  3.1.1 Open data requirements ...................................................................... 68
  3.1.2 The legal status of open data .............................................................. 70
3.2 Open data: some historical and genealogical notes .................................... 73
  3.2.1 Europe ................................................................................................ 74
  3.2.2 International developments ................................................................. 78
3.3 Open data legislation in the EU ................................................................. 82
  3.3.1 The PSI Directive ................................................................................ 82
  3.3.1.1 History .......................................................................................... 82
  3.3.1.2 Structure ....................................................................................... 85
  3.3.1.3 The recast of the PSI Directive ..................................................... 87
  3.3.1.4 Incentives and past behaviour: a peek at the EU data economy .... 94
  3.3.2 The INSPIRE Directive ...................................................................... 96
  3.3.2.1 History .......................................................................................... 96
1. Introduction

This dissertation deals with the idea of balancing open data and data protection, as their apparent conflict, it is argued, may hamper the development of the so-called ‘smart city’: the facet of the digital revolution that interacts with the built environment. The incompatibility between open data, which is postulated to be a requirement for cities to be ‘smart’, and data protection legislation, may render urban intelligence hard to achieve.

The research questions from which the thesis moves from originate within the ‘Safeguarding data Protection in an Open data World’ (SPOW) project.¹ They maintain that EU data protection legislation – its material scope in particular – would be overly expanding, and that the right to personal data protection would be exceedingly straying away from its original framing: the right to privacy. The SPOW project thus postulated that something ought to be done, in terms of balancing open data and data protection, lest we forego the promises of the cities of the future.

This chapter introduces the dissertation. It is structured as follows: the next section introduces the context of the research – smart cities, open data, and data protection. The following section lists the thesis’ research questions, and their significance. The third section discusses this thesis’ methodology, and the fourth and closing section outlines the chapters to come.

1.1 Smart cities, open data, and data protection

It can be argued that we are on the verge of a revolution in urbanism – a shift from data-informed urbanism to data-driven,² networked urbanism.³ An ever-increasing deluge of data⁴ is being collected, analysed, and used to fuel what has been commonly defined with the umbrella term “smart city”: an environment in which an extended network of sensors, coupled with big (and not-so-big) data analytics techniques, produce an extremely large amount of data – often in real time – allowing to manage and control diverse facets of the urban ecosystem, with a higher level of responsiveness and more targeted and granular options. Smart cities are significantly based upon their sensing capabilities and, through their vast sensors’ network and

¹ NWO/STW Maps4Society program (project number 13718).
² “In order for cities to perform well on the above dimensions, for improvement there is a need for evidence-based planning, which will enable a better identification of problematic sectors (e.g. transport) and areas (e.g. neighbourhoods) and a better allocation of resources. Such evidence-based planning is in desperate need of analytics and of relevant data to reveal caveats at a fine-grained scale, both in terms of space and time”: John Steenbruggen, Emmanouil Transos and Peter Nijkamp, ‘Data from Mobile Phone Operators: A Tool for Smarter Cities?’ (2015) 39 Telecommunications Policy 337.
⁴ Already in 2012, it has been claimed that the amount of data mankind produced in two days – at the time, five exabytes of data – is roughly equivalent to all the data produced between the beginning of civilization and the year 2003: see Rob Kitchin, The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences (Sage 2014) XV. See also Christopher Kuner and others, ‘The Challenge of “Big Data” for Data Protection’ (2012) 2 International Data Privacy Law 47: "The Economist reports in its 2012 Outlook that the quantity of global digital data expanded from 130 exabytes in 2005 to 1,227 in 2010, and is predicted to rise to 7,910 exabytes in 2015. An exabyte is a quintillion bytes. If you find that hard to visualize, consider this: someone has calculated that if you loaded an exabyte of data on to DVDs in slimline jewel cases, and then loaded them into Boeing 747 aircraft, it would take 13,513 planes to transport one exabyte of data. Using DVDs to move the data collected globally in 2010 would require a fleet of more than 16 million jumbo jets".
through all the devices interacting with it, feed back into the model the information gathered. Thus, the urban built environment is morphing in two distinct yet parallel directions. On one hand, its development is getting data-driven, rather than merely data-informed: information is directly shaping (adaptive) architecture, rather than being only a tool available for that purpose. On the other hand, the (smart) city environment itself functions as a data-gathering infrastructure, which enables the collection of an unprecedented quantity of data, to be subsequently used in decision-making processes, and potentially shared and reused for additional value. The general sentiment is that the cities of the future — smart cities, intelligent cities — will be built on data as much as present-day ones are built on land.

The benefits deriving from the improvements in urban governance that smart cities are meant to bring along are undeniable, and often intuitive: public transportation infrastructures built where they are needed the most, services deployed in the areas most accessible for the population segment to which they are targeted, better resource allocation, a more participatory and inclusive governance, based on the fact that the decisions that shape the environment we live in are taken considering the aggregated data gathered by the city itself, and so forth. As of now, “smart city” is mostly a buzzword, a multifaceted concept whose characteristics have yet to be precisely outlined, and that therefore lends itself to quite a degree of definitional confusion. For the purposes of this introductory chapter, we will initially define the smart city as a city in which information and communication technologies (ICT) are intertwined with the urban environment, enabling, coordinating or integrating the functioning of its infrastructures; in the smart city, ICT is a precondition for the urban environment to function as intended. In this preliminary definition, ICT is meant to encompass both the hardware layer — a networked array of sensors, actuators and instruments that will automate part of the city’s functions — and the software one.

The smart city’s development should not however be tout court equated with the development of the technology stack underlying it. As it has been noted, smart city literature has been in a widespread agreement about the fundamental importance of the human factor in the

6 See Anthony M Townsend, Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia (WW Norton & Company 2013).
9 See Annalisa Cocchia, ‘Smart and Digital City: A Systematic Literature Review’ in Renata Paola Dameri and Camille Rosenthal-Sabroux (eds), Smart City (Springer 2014).
10 In this sense, the smart city constitutes a code/space: “[a]ny space that is dependant on software-driven technologies to function as intended”, according to the definition given by Rob Kitchin and Martin Dodge, Code/Space: Software and Everyday Life (MIT Press 2011). See also M Batty and others, ‘Smart Cities of the Future’ (2012) 214 The European Physical Journal Special Topics 482: “Smart cities are often pictured as constellations of instruments across many scales that are connected through multiple networks which provide continuous data regarding the movements of people and materials in terms of the flow of decisions about the physical and social form of the city”.
11 See Neirotti and others (n 8); Batty and others (n 10).
smart city’s conceptualization, development and deployment. The deployment of the underlying ICT infrastructure and of the big data analytics13 (BDA) techniques required for large-scale smart environments to function as intended indeed necessitates, in parallel, of both a substantial investment in human and economic capital and of an adaptation of urban practices, conditions and governance. In this respect, it appears particularly worth noticing so far two distinct yet complementary approaches appear to be prevalent in the contextualization of the smart city environment: a “top-down” approach, “closely related to the technologically deterministic idea of a "control room" for the city”,14 and a bottom-up one,15 in which the focus is not on the technologies on which the concept of smart city depends, but rather on the human aspect of smart environments – on smart citizens, rather than on smart infrastructures, and on new modalities of urban governance, rather than on mere technical advancement.

Central to this preliminary definition is the fact that the ICT on which the smart city is based will produce an enormous, constant stream of data pertaining to the city’s physical and social structure, which is meant to be analysed and synthetized for a number of purposes, amongst which further urban development looms large.16 The reasons for which the information gathered through and by the smart city are to be used are manifold, but ultimately ideally directed towards sustainable economic advance, increases in the population’s quality of life, improvements in the management of natural resources, and stimulation of participation and inclusiveness in the urban environment’s governance.17

The constant availability and accessibility of the unprecedented amount of information that smart cities are bound to bring forth, however, warrants a cautious approach, and calls for clear-cut values in order to orient the design of the data gathering and processing infrastructures on which smart cities will be based. Indeed, on one hand, the data gathered by and through the smart city environment can revolutionize urbanism, and therefore enable a plethora of positive effects and constructive consequences. On the other hand, the array of networked sensors and the extensive data processing capabilities that define the smart city’s


15 “There are some people who are thinking in a top-down way, putting a lot of new sensors into the city,” says Carlo Ratti, director of the Senseable City Lab at the Massachusetts Institute of Technology. Singapore is a leading example. “Or you can also look at a more bottom-up, distributed way where you can use what you already have, such as a cellphone,” Mr Ratti adds.” – Tim Bradshaw, ‘Mobiles Could Be the Secret to “Smart” Cities’ Financial Times (22 February 2016). See also Cocchia (n 9); Neirotti and others (n 8); Anthony M Townsend, ‘Life in the Real-Time City: Mobile Telephones and Urban Metabolism’ (2000) 7 Journal of urban technology 85.

16 For instance, mobile phones’ data (e.g. Call Data Records) are increasingly been seen as an information source to be used for urban planning and development, rather than the mere by-product of a communication tool, since they allow to better understand and model human activities – a necessary step to be undertaken to understand urban dynamics; see Steenbruggen, Transos and Nijkamp (n 2). On the importance of mobile computing in mapping the urban environment, see also Nathan Eagle and Kate Greene, Reality Mining: Using Big Data to Engineer a Better World (MIT Press 2014).

17 In this sense, a fitting definition has been given by Andrea Caragliu, Chiara Del Bo and Peter Nijkamp, ‘Smart Cities in Europe’ (2011) 18 Journal of urban technology 65. According to the cited paper, smart cities are environments in which “investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance".
technological stack raise a number of legal and policy issues, which need to be tackled from the very outset of the smart city’s development\textsuperscript{18} – from the design phase on.

Privacy and data protection,\textsuperscript{19} in primis, are naturally threatened by the deluge of data gathered by the multiplicity of sensors on which the smart city is based, which can be stored, processed, and analysed to potentially allow, \textit{inter alia}, the identification of whoever resides in the city, the inference of his or her personal characteristics, and the application of a profile (or a ‘shadow profile’)\textsuperscript{20} based on the latter. The future evolution of large-scale smart environments has the potential to shift the normality of urban dwelling from a paradigm in which anonymity is the norm and identification the exception to one in which inhabitants are identified by default, and anonymous by exception, thus undesirably shifting the balance between city residents and local authorities.\textsuperscript{21} Moreover, networked objects are vulnerable to attacks by malicious intruders and to accidental malfunctions, which could compromise the smart city’s security, and hence the confidentiality\textsuperscript{22} of the information processed in and by it. Finally, algorithmic decision-making processes and big data analytics have the potential to enhance possible discrimination and exclusion\textsuperscript{23} from the very same urban decisional processes that the smart city ideal would aim at making more inclusive and accessible.\textsuperscript{24} Some of the potential threats to individuals’ privacy and data protection rights are evident \textit{ictu oculi} even when adopting an initially vague definition of what smart cities and large-scale smart environments are, and without yet specifically considering the individual technologies likely to be deployed in them.

Yet, while the data avalanche characterizing the ICT era is indeed worrisome from a privacy and data protection perspective, it is also a formidable driver for growth and innovation.\textsuperscript{25} Data

\footnotesize
\textsuperscript{18} Rectius, from the very outset of the deployment of the technologies and policies that have the potential effect of transporting a traditional city into a ‘smart’ one – which is a diachronic process, rather than a synchronic implementation.

\textsuperscript{19} This thesis’ conceptualization of the rights to privacy and data protection reflects the one dominant in the EU milieu, according to which they are two separate – yet undeniably closely connected – fundamental rights: see articles 7 and 8 of the Charter of Fundamental Rights of the European Union (2012/C 326/02). On the distinction between privacy and data protection as fundamental rights, see Raphaël Gellert and Serge Gutwirth, ‘The Legal Construction of Privacy and Data Protection’ (2013) 29 Computer Law & Security Review 522. On the emergence of data protection as a standalone fundamental right, see Gloria González Fuster, \textit{The Emergence of Personal Data Protection as a Fundamental Right of the EU} (Springer Law, Governance and Technology Series 2014); Orla Lynskey, ‘Deconstructing Data Protection: The “Added-Value” of a Right to Data Protection in the EU Legal Order’ (2014) 63 International and Comparative Law Quarterly 569; Orla Lynskey, \textit{The Foundations of EU Data Protection Law} (Oxford University Press 2015).

\textsuperscript{20} i.e. a profile based on the characteristics of the individual’s ‘closest neighbour’, rather than on the characteristics of the individual itself.


\textsuperscript{22} The CIA (confidentiality, integrity and availability) triad is a commonly used definition of the elements composing the overarching concept of security.


\textsuperscript{24} Caragliu, Del Bo and Nijkamp (n 17).

\textsuperscript{25} On the topic, see Eagle and Greene (n 16).
is often defined as “the new oil”\textsuperscript{26} – a metaphor that seems to resonate with many,\textsuperscript{27} but that does not however render full justice to the promises and perils of information (re)use. Rather than a perishable commodity, data is more like an infrastructure: once available, it has the potential to keep generating value; to continue with the imagery of the metaphor above, data can be better portrayed as a pipeline through which insights, knowledge and understanding – the actual oil – flow.

The value deriving from the reuse\textsuperscript{28} of public sector information (PSI) alone – therefore excluding data generated by public-private partnerships (PPPs) and by the private sector – is hardly precisely quantifiable,\textsuperscript{29} yet the available estimations confirm the idea that the secondary use of data keeps generating an impressive amount of worth. A study conducted in 2006 “estimates for the overall market size for public sector information in the European Union range from €10 to €48 billion, with a mean value around €27 billion. This amounts to 0.25% of the total aggregated GDP for the European Union and Norway (€10.730 billion)”.\textsuperscript{30} In 2015, another study estimated that “the direct market size of Open Data is expected to be 55.3 bn EUR for the EU 28+. Between 2016 and 2020, the market size is expected to increase by 36.9%, to a value of 75.7 bn EUR in 2020. The total market value of Open Data is estimated between 193 bn EUR and 209 bn EUR for 2016 with an estimated projection of 265-286 bn EUR for 2020, including inflation corrections. For the period 2016-2020, the cumulative direct market size is estimated at 325 bn EUR. The cumulative total market size for Open Data is forecasted to be between 1,138 and 1,229 bn EUR”\textsuperscript{31}. From both an economic and a social perspective, open data undoubtedly generates a remarkable amount of seemingly ever-increasing value and growth. Smart cities, through their data gathering and computing capabilities, are hence expected to provide a large boost to the amount and granularity of the information we are able to collect and process, hence fostering social and technological innovation and economic progress.

There is indeed, undeniably, a growing tendency to release all sorts of data through the Internet. Both national and local administrations are opening up their databases and allowing individuals, companies and other administrations to freely re-use public sector information without any sort of restrictions in their usage – a stance which is often referred to as ‘open


\textsuperscript{28} I.e. the use of documents, or parts thereof, by either natural or legal persons, irrespective of the medium in which they are contained, for purposes that differ from the ones within the (public) task for which they were initially produced.

\textsuperscript{29} “\textit{While there is no intrinsic value in open data, the benefits are a result of value-added processes}” – Maureen Henninger, ‘The Value and Challenges of Public Sector Information’ (2013) 5 Cosmopolitan Civil Societies: An Interdisciplinary Journal 75, 85.


data’. Open data can be initially defined, for the purposes of this chapter, as data that “is free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness”: see Open Knowledge International, ‘Open Definition 2.1’ <http://opendefinition.org/od/2.1/en/> accessed 22 January 2016. Open data, in a nutshell, refers to information made available as a whole and at no cost, in a convenient, modifiable, interoperable and machine-readable form, under terms that permit others to use, re-use and redistribute it, and to merge or cross-correlate it with other datasets. Everyone has to be able to do so: there has to be no discrimination or restriction of any sort against ventures or behaviours, or against persons or groups.

In the EU, the re-use of PSI is advocated through diverse legislative initiatives, such as the Directive on the re-use of public sector information (PSI Directive), which aims at maximizing the re-use of public sector information, or regarding geographic data — the INSPIRE Directive, which aims at creating an EU-wide geographic data infrastructure allowing a wide array of subjects to discover, view and freely reuse geographic datasets. Moreover, European institutions, as well as many European Member States, are increasingly promoting and implementing open data policies and practices. The underlying hope is that the greater availability of interoperable public data catalyses secondary use of such data, which would then lead to scientific and economic growth, better government transparency and a more inclusive governance. However, it needs to be considered that open data is not an absolute value: it needs to be reconciled with all the other rights and interests that may be dented by the use or abuse of the information released, amongst which the rights to privacy and data protection are arguably paramount. The rise of the smart city and of the informational deluge it is bound to bring along will arguably render the open data vs. data protection debate even more timely, actual, and interesting.

The interpretation and application of privacy and data protection legislation is therefore tricky when considered in light of the possible evolution of future large-scale smart environments: on one hand, their potential for surveillance, control, and intrusion upon city dwellers’ personal

32 Open data can be initially defined, for the purposes of this chapter, as data that “is free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness”: see Open Knowledge International, ‘Open Definition 2.1’ <http://opendefinition.org/od/2.1/en/> accessed 22 January 2016. Open data, in a nutshell, refers to information made available as a whole and at no cost, in a convenient, modifiable, interoperable and machine-readable form, under terms that permit others to use, re-use and redistribute it, and to merge or cross-correlate it with other datasets. Everyone has to be able to do so: there has to be no discrimination or restriction of any sort against ventures or behaviours, or against persons or groups.

33 See e.g. Walravens, Breuer and Ballon (n 14); Open North, ‘Open Smart Cities Guide V1.0’ (2018).


36 As it has been noted, “there may be an emerging shift in these processes from a representative government to participative or direct government because of new and powerful information and communication technologies” – Henninger (n 29) 81.

37 Rectius, the rights and freedoms that can be upheld through open data sharing and re-use are not absolute values.
sphere warrants a careful scrutiny of the implementation of the underlying technologies. On the other, both the smart city’s direct goals of fostering economic growth, inclusive governance, and optimal resource allocation, and the value and opportunities granted by the information reuse, suggest a cautious approach in determining if and to what extent personal data protection legislation applies to the particular case in point. The very definition of what can be considered as personal data (“any information relating to an identified or identifiable natural person […] one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity”), for instance, is oftentimes foggy, and subject to a wide degree of interpretative discretion that contributes in blurring the boundaries between personal and non-personal information, and thus to legal uncertainty.

A balance between open data instances and data protection requirements, the SPOW project argued, is therefore needed, as an overly extensive interpretation of personal data protection legislation may hinder innovation and economic development, while at the same time the uncontrolled availability of public datasets will lead to widespread profiling and surveillance activities, bringing forth chilling effects on individual freedom and a reduction in democratic accountability.

1.2 RESEARCH QUESTIONS AND SIGNIFICANCE

The central aim of this thesis is to investigate the balance between open data and data protection in the specific context of future smart city ecosystems, and to consequently develop a set of related recommendations and best practices. The main research question, therefore, is: “how should the right to data protection be balanced with the interests underlying open data regulations in order to facilitate the realization of the smart city ecosystem?”. In order to answer this central issue, a number of additional research sub-questions have been drafted:

---

40 See Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (2007). According to the A29WP, for instance, “in order to consider that the data “relate” to an individual, a “content” element OR a “purpose” element OR a “result” element should be present” – the A29WP’s interpretation of the concept of personal data, while underlying the fact that it is a plastic and technologically neutral notion, allows to point out how the definition can be oftentimes uncertain, and always dependent on the context in which the processing activities take place.
43 See e.g. Elizabeth Stoycheff, ‘Under Surveillance: Examining Facebook’s Spiral of Silence Effects in the Wake of NSA Internet Monitoring’ (2016) 1 Journalism & Mass Communication Quarterly 16.
1. How can open data regulations, in the light of technological advances and the greater availability of (geographic) data, be problematic from the perspective of the right to the protection of personal data?

2. To what extent is EU data protection law in conflict with open data requirements?

3. To what extent are the conflicting regulations a barrier to the full development of the concept of smart city? How should open data and data protection be balanced?

4. Should data protection be brought back to its original ‘privacy dimension’?

The relatively scarce privacy and data protection-related literature concerning smart cities has mainly focused on practices such as surveillance, monitoring, ‘dataveillance’, and predictive analytics. This thesis aims at investigating the data protection implications of publicly sharing large sets of (public sector) data, and will provide insights relevant to both the open data community and to the data protection debate.

As mentioned, open data has potentially a great commercial and societal potential, and is fuel to the development of smart cities. However, this potential and development can only be reached if the citizens’ rights to privacy and data protection are adequately safeguarded. EU data protection law can potentially be a significant obstacle for both the smart city’s development and the success of open data initiatives; at the same time, future large-scale smart environment and the ever-increasing and unprecedented amount and availability of data are naturally bound to be threatening in respect to individuals’ privacy and data protection rights. The research aims at giving relevant stakeholders guidance in the application of open data and data protection regulation, clarifying both landscapes and supporting an optimal development of the framework of values on which the design and development of the smart city environment should be based.

The research is also very timely. When I began writing this dissertation, the process of reform of the European data protection framework, which started in January 2012 with the Commission’s proposal for a General Data Protection Regulation (GDPR), was just starting to see its end: the 15th of December 2015 the representatives of the European Parliament, the Council and the Commission (participating in the so-called ‘trilogue’, informal meetings attended by delegates of the abovementioned institutions) reached a political agreement on the text of the forthcoming GDPR. The final text of the regulation entered into force in 2016, and became applicable two years later, in 2018.50

Likewise, the writing of this thesis began at the end of 2015, when the PSI legislation in force was the 2003 PSI Directive as amended in 2013, and ended in 2019, after its scheduled review.

---

45 See e.g. Finch and Tene (n 21); Edwards (n 7).
50 GDPR, Art. 99.
51 See Art. 13 of the 2013 PSI Directive.
led to the drafting and subsequent adoption of its recast. The timespan during which the research took place allowed the examination of both the GDPR (which is expected to have a deep impact\(^{52}\) on global data protection practices) and of the 2019 recast of the PSI Directive from their very genesis to the first period of application after their formal adoption.

The same is also true for the development of the concept of smart city, which seems, as of now, still very foggy, and for its relationship with both the data protection regulatory framework, on one hand, and with open data (and PSI reuse) instances, on the other. Modern cities are evolving, and their transformation into large-scale smart environments will undoubtedly intersect with several individual and collective rights and interests, such as the right to the protection of personal data, or the legitimate interest to the release of public sector information in an open format. Such an interaction warrants preemptive scrutiny, and the current level of development of the concept of smart city appears to suggest the timeliness of the research undertaken by this thesis.

Furthermore, the current amount of legal literature regarding the balance between open data and data protection\(^{53}\) would certainly benefit from a more extensive and in-depth analysis of the issue at stake; current research, moreover, tends to look at the existing urban landscape rather than at how it is going to evolve in the near future, e.g. how PSI reuse and open data disclosures will be shaped by the advent of large scale smart environments. It is still quite soon – both chronologically and with respect to this thesis’ progress – to pinpoint how smart cities will shape the relationship between open data and data protection. However, a number of clues\(^{54}\) hint at the possibly enormous boost that smart cities initiatives could get from – and give to – the reuse of PSI, and to the open data movement in general. It appears intuitively significant to examine whether and how this impacts the clash naturally running between open data interests on one hand and individuals’ rights to privacy and data protection on the other.

Finally, we are arguably just at the dawn of the transition from traditional cities to smart cities; as it has been noted,\(^{55}\) the concept of smart city is full of both promises and potential dangers, and its development would certainly benefit from the identification of a set of legal principles and guidelines through which to design the information infrastructures underlying it. Moreover, while there is plenty of literature defining the privacy and data protection aspects of the Internet of Things (IoT), it is still as of now unclear if the difference in scale between the IoT (a small-scale smart environment)\(^{56}\) and the smart city (a large-scale smart environment) leads to different problems and concerns, or if the underlying problematics – and the ways to tackle them – remain unchanged.

This research will therefore provide insights on how to reconcile the ever-growing need for data availability, which is nowadays at least partly satisfied through open data sharing and reuse, with the rights to privacy and to the protection of personal data of individual data

---

53 E.g. Zuiderveen Borgesius, Van Eechoud and Gray (n 41); Kulk and Van Loenen (n 39); Cristina Dos Santos and others, ‘LAPSI Policy Recommendation n. 4 Privacy and Personal Data Protection’; Cristina Dos Santos, ‘On Privacy and Personal Data Protection as Regards Re-Use of Public Sector Information (PSI)’ (2012) 6 Masaryk UJL \& Tech. 337.
54 See Walravens, Breuer and Ballon (n 14).
56 For a disambiguation between the concepts of large-scale and small-scale smart environments, see the next chapter.
subjects, which can potentially be threatened by the introduction of large-scale smart environments.

1.3 METHODOLOGY

This thesis, albeit aiming at developing a holistic and interdisciplinary approach, is prevalently a work of legal scholarship, albeit embedded in the broader and multidisciplinary field of the research for the built environment. Hence, it is bound to having to deal with a common preconception regarding legal research: “(t)he predominant view of lawyers is that they are not really academic--"arcane, distant and alien: an appendage to the academic world". Their personal qualities are dubious: vociferous, untrustworthy, immoral, narrow, and arrogant: though kinder eyes see them as impressive and intelligent. The discipline is variously described as unexciting, uncreative, and comprising a series of intellectual puzzles scattered among "large areas of description".

Indeed, legal research is a relatively new addition to the disciplines comprising the milieu of the research for the built environment, and its methodologies (rectius, the prevalent lack thereof) have been generally frowned upon by both natural and social science practitioners. This does not come as a surprise, considering the differences running between the law and the other disciplines comprising the built environment’s field of research: scholars in architecture, urbanism, economics and management, for instance, mostly rely on empirical qualitative and/or qualitative data in order to develop, test and validate their hypothesis, and thus the validity of their inquiry directly depends on the validity of their empirical research methodology. Legal research, in contrast, can be led back to the humanities’ research’s characteristics – subjective, argument-driven, and based on others’ authoritative opinion and underlying reasoning rather than on the scientific method.

57 “The built environment is usually considered to be an interdisciplinary (or, at the very least, a multidisciplinary) field linking the disciplines of management, economics, law, technology and design”: Paul Chynoweth, ‘Legal Research’ in Andrew Knight and Les Ruddock (eds), Advanced research methods in the built environment (John Wiley & Sons 2009) 28.

58 “Legal researchers have always struggled to explain the nature of their activities to colleagues in other disciplines” – Chynoweth (n 57).

59 Tony Becher, ‘Towards a Definition of Disciplinary Cultures’ (1981) 6 Studies in Higher Education 109. Van Hoecke stated that “the criticism of legal doctrine is partly founded: it is often too descriptive, too autoepoietic, without taking the context of the law sufficiently into account; it lacks a clear methodology and the methods of legal doctrine seem to be identical to those of legal practice; it is too parochial, limited to very small scientific communities, because of specialisation and geographical limits; there is not much difference between publications of legal practitioners and of legal scholars. All this may be correct, but as such it does not disqualify legal doctrine as a discipline in its own right, with its own, appropriate, methods”: Mark Van Hoecke, ‘Legal Doctrine: Which Method(s) for What Kind of Discipline?’ in Mark Van Hoecke (ed), Methodologies of legal research: which kind of method for what kind of discipline? (Hart Publishing 2011).


61 It appears interesting to mention, in this regard, that “(d)uring the whole of the middle-ages, legal doctrine was highly thought of and considered as a ‘scientific discipline’, as in those times ‘authoritative interpretation’, not ‘empirical research’, was the main criterion for the scientific status of a discipline. Slowly as from the seventeenth century, but mainly as from the nineteenth century, this changed dramatically. The success of the positive sciences altered the conception of ‘science’ in western societies. Physics became the model. Hence, a combination of empirical data, mathematics, testing of hypotheses, developing theories with a general validity and without geographical limitations, became the ideal for any ‘scholarly discipline’”: Van Hoecke (n 59) 1.
Doctrinal research has customarily been the prime method through which legal research has been performed. The formulation of legal doctrines – i.e. the systematic and normative evaluation and interpretation of legal norms and case law made by a qualified interpreter – is the most traditional line of legal scholarship reasoning. It aims at defining what is the applicable law in a particular context, putting it in a specific logical and hierarchical structure, and identifying (and ideally solving or clarifying) potential inconsistencies and incongruences.\(^{63}\)

Doctrinal research, in a nutshell, investigates what the applicable law is, and whether it applies in the hypothesis in consideration, mostly using the letter of the law, the relevant courts’ cases (i.e. jurisprudence), and other scholars’ and professionals’ doctrinal output (i.e. their own authoritative interpretation of the law). As a consequence of its reliance on authoritative interpretation, rather than on scientific and empirical methods, the worth of traditional doctrinal research largely depends on the extent of the consensus of the legal scholastic community, rather than on quantitative or qualitative empirical evidence, as it is the case for natural and social sciences and as opposed to the humanities’ sectors.

However, all doctrinal analysis but the purest can hardly be completely disjointed from the context from which the legal inquiry takes its moves. Perhaps with the exception of pure legal philosophy, there is barely any kind of doctrinal study that can leave aside e.g. the historical, political or social context in which the analysis takes place, or the technological or organizational characteristics of the environment disciplined by the legal norms analysed by the scholar. In particular, the issues this research aims at tackling are largely deriving from the societal impact of rapid technological developments, and it would therefore seem both sterile and redundant to perform purely doctrinal legal research (i.e. merely questioning “what is the law?”), without considering both the technological drivers and the social implications of the smart city and of its enablers.

This thesis’ legal analysis style therefore leans towards an interdisciplinary approach, thus inquiring about the law as a social construct,\(^ {64}\) rather than just researching in the law, as traditional doctrinal analysis does. This approach appears to be more suitable both to the technology-driven topic of this thesis’ legal inquiry, and to the milieu in which this study is framed – the research for the built environment. Moreover, one of the aims underlying the research project from which this thesis stems is the creation and development of a set of


\(^{63}\) Even though, as it appears necessary to point out, doctrinal studies are far from being a unitary form of research, and can assume quite a wide array of diverse characteristics: on a case by case basis, its nature could be e.g. explanatory, empirical, hermeneutic, explorative, logical, instrumental, or evaluative; moreover, it could be supported by diverse meta-juridical disciplines (e.g. legal history, legal sociology, legal anthropology; legal psychology, law and biology, law and economics, etc.); see Mark Van Hoecke, ‘Preface’ in Mark Van Hoecke (ed), *Methodologies of legal research: which kind of method for what kind of discipline?* (Hart Publishing 2011) V.

\(^{64}\) As it has been noted, ‘we have seen, as from the end of the nineteenth century, and mainly in the course of the twentieth century, the birth and development of other social sciences focusing on law: legal sociology, legal psychology, law and economics. all of those disciplines offer empirical research and theory building in legal matters. However, they never aimed at replacing legal doctrine, but just wanted to supply legal scholars, legal practitioners and policymakers with useful information on legal reality [...] so, today, there is a somewhat schizophrenic situation in which one discipline, legal doctrine, is basically studying law as a normative system, limiting its ‘empirical data’ to legal texts and court decisions, whereas other disciplines study legal reality, law as it is’; see Van Hoecke (n 59) 2.
recommendation aspiring at facilitating practitioners, academics and policymakers dealing with the intersection between open data and data protection: an interdisciplinary constituency appears therefore more fitting than a traditional, purely doctrinal one.

When reading this thesis, it may be useful to consider that the SPOW project began at the end of 2015, and the research proposal it originates from was written well before then. The research proposal, and the research question substantiating it, moreover, contain a set of assumptions that are hardly uncontroversial. I am writing this paragraph in late 2019, and it seems important to report on how the interpretation of the research questions, and the methodology used to answer them, changed (at least partly) as the research progressed.

A first reason for this shift derives from a heightened understanding of the matter(s) at stake. As the following chapter explain, the research questions assume a different dimension when one considers what open data is from a strictly normative perspective, what the right to personal data protection has become at a (quasi)constitutional level, and how the secondary legislation through which that right substantiates actually functions.

A second reason is due to an ever-changing regulatory landscape which, for once, did not actually change that much, and to the causes of that. This thesis began when the GDPR was still a legislative proposal, and its content was still being determined, and when the 2013 PSI Directive had yet to undergo the formal evaluation that would eventually lead to its 2019 recast. The end result of the reform processes that reshaped both data protection and PSI law did not alter the pre-existing relationship between PSI and data protection, which is, by now, an established part of EU law. The reasons for this choice – arguably the most reasonable one amongst the available options – turned out to be particularly interesting.

Beyond the qualification of the elements that compose the research question, and the analysis of the (lack of) reform of the applicable law, a third reason for the shift in the original research design derives from reading between the lines of the research questions. The space between the elements that are made explicit by the research questions, in other words, ended up not being empty, but rather more significant than said elements themselves.

The research approach followed during the development of this thesis was conditioned by a number of early findings, which either resulted contrary to some of the research assumptions, or that highlighted topics of paramount importance that were neglected in the initial design. A first amendment concerns the smart city and its requirements, or rather the idea that it is a unitary concept that has requirements in the first place. It seemed thus important, through a critical analysis of the available smart city literature and of the relative discourse, to figure out what is generally meant with the wording ‘smart city’. That, in turn, compressed the role that the case studies initially planned had in the research carried out.

A second deviation from the original design concerns the fact that the idea of balancing, perhaps counterintuitively, is not as neutral as the original research proposal suggested. As the following chapters will discuss, the idea of balancing open data and data protection is bound to substantiate itself in a compression of the right to data protection, which is to be assessed through proportionality testing. ‘Balancing’ has a precise and technical meaning in EU law, and

---

65 Albeit I cannot claim methodological soundness, since (albeit curious about other disciplines by nature) I am a legal scholar by training, my analysis was partly informed by discourse analysis: see e.g. Deborah Tannen, Heidi Ehernberger Hamilton and Deborah Schiffrin, The Handbook of Discourse Analysis (Wiley Online Library 2015); Teun A van Dijk, ‘Principles of Critical Discourse Analysis’ (1993) 4 Discourse & Society 249; Robert Chia, ‘Discourse Analysis Organizational Analysis’ (2000) 7 Organization 513.
is expressed through the concept of proportionality, which provided for the conceptual backbone for the discussion of the relationship between open data and personal data protection. The investigation of meaning and mechanisms of balancing as *lato sensu* proportionality thus took a prominent place in the research design.

A third deviation from the original research proposal concerned the plan of drawing lessons from the US, Canada, and the UK. While there are a number of documents and doctrinal sources that have been used as a support for this research that come from North American jurisdictions, it seems that EU data protection law and policy (thus including the UK’s) have a higher degree of maturity than their North American counterparts,⁶⁶ which reduces the possibility of drawing ‘lessons’ from how those countries deal with the balance between open data and data protection. Moreover, from a more practical perspective, it resulted clear that the ‘best practices’ used in continental Europe and across the English Channel are extremely similar to the ones available across the Atlantic, irrespective of regulatory differences.

Traditional legal research thus maintained a preponderant importance, remaining the central method of inquiry used in the development of this study; accordingly, the main sources to be found in this thesis’ writing are legislation, case law, and authoritative legal doctrine. However, the law is considered as part of a broader set of regulatory forces,⁶⁷ whose interplay provides a more lifelike reading on the interests to be safeguarded in the development of the smart city environment. The partially practical constituency which this thesis aimed at achieving also suggested employing (unstructured) qualitative interview methods⁶⁸ to get real-world insights while performing research in an academic setting. The participants were selected mostly – albeit not exclusively – amongst the members of the user committee of the project in which this study is embedded.

The project within which this research has been conducted benefitted from the presence of an extended and diverse user committee, comprising members belonging to both the private and to the public sector (SMEs and larger, multinational companies; municipal and national agencies). The presence of such user committee allowed for this academic research to draw from the perspective of practitioners operating at different levels, and responding to different stakeholders and priorities. The dissimilar and often conflicting viewpoints and perspectives that practitioners have when compared to academics suggested a co-research⁶⁹ approach, a useful tool to better understand complex social phenomena such as the development of smart cities, the growth of open data sharing and reuse, and their effect on individuals’ rights to privacy and data protection. Accordingly, in the development of this study, the perspective of

---


practitioners working at different levels and at different scales in both the public and the private sector have been sought for.

1.4 **OUTLINE**

After this brief introduction, the following chapter will introduce the notion of ‘smart city’ and the characteristics of large-scale smart environments, their enablers, and their relationship with the two overarching topics of this thesis: data protection, on one hand, and open data, on the other. The next chapter aims at outlining the thesis’ understanding of what the wording ‘smart city’ means, whether it is possible to enucleate its requirements, and whether the achievement of those requirements is hampered by the conflict of open data disclosure policies and the right to the protection of personal data. It argues that, at least for the purposes of this thesis, the smart city is to be considered as a narrative, albeit supported by a technological underpinning, and its emergence should not change the equilibrium between open data and data protection. Ultimately, what really counts are the values driving the instrumentation of the built environment, the technologies used, and the consequences of their deployment; everything else serves, I believe, needs of marketing and obfuscation.

The third and fourth chapters are about open data and data protection, respectively. The third chapter clarifies what open data is from both a practical perspective and from a normative one. The chapter also distinguishes it from PSI, accounts for the latter’s evolution in EU law, and for (some of) the other legislative drivers of open data in the Union. The fourth chapter revolves around the fundamental right to personal data protection. It deals with its emergence, and with its substance and essence at a (quasi)constitutional level. The chapter also explores data protection’s formal and substantial distinction from the right to privacy, underlining the anachronisms of the idea of bringing data protection back to its privacy roots.

Having dealt with the three main building blocks of this thesis – smart cities, open data, and data protection – the fifth chapter explores the meaning of balancing and (thus) of proportionality. It clarifies how the idea of balancing open data and data protection entails an unnecessary deregulation of personal data protection law, as personal data does not necessarily have to be kept ‘closed’, but can rather exist within several shades of the openness spectrum, of which open data is an extreme end. The chapter then discusses the idea of performing a proportionality test to a hypothetical measure able to compress the right to personal data protection up to the point where personal data could be shared as open data without diluting its meaning and the degree of openness it requires. It highlights how open data should not be considered as a value to be balanced with personal data protection – an input of a *lato sensu* proportionality assessment – but rather as a potential output of the balancing test between the *rights and freedoms* that may be supported through open data sharing and re-use and any right that may compete, such as data protection.

The sixth chapter explores what I came to believe to be the real core of the matter: misunderstandings about the notion of personal data and the material scope of EU data protection law. The SPOW project was originally motivated by legitimate concerns about the over-inflation of the concept of personal data, and about its consequences upon kinds of information traditionally deemed safe for publication as open data. The idea of having to

---

70 E.g. mobile computing, cloud infrastructures, ubiquitous computing, the Internet of Things (IoT), (big and small) data analytics, etc.
somehow ‘balance’ open data and data protection, and the reference to the ‘smart city’ as its justification, seem to be, ultimately, a sectorial\textsuperscript{71} manifestation of dissatisfaction about the trade-off between the legal certainty of the material scope of EU data protection law and the kind of flexible and contextual protection it is meant to grant.

The seventh chapter ends the thesis. In it, conclusions are drawn, limitations acknowledged, and directions for further research given.

---

\textsuperscript{71} As it may be representative of the concerns of (part of) the open data community, but it does not seem to be the expression of a broader consensus.
2. **On the Concept of ‘Smart City’**

The digitalization and networking of the urban environment\(^{72}\) that has been happening in cities all over the world for the past few decades is a complex phenomenon, which concerns all the layers and facets of a city’s social fabric. As modern Information and Communication Technology (ICT) intertwines with the urban landscape, several aspects of the traditional city are changed, sometimes entirely revolutionized. Urban transportation, for instance, has historically been a field where technological evolution has had a direct impact on cities’ development and growth. ICTs – whether directly embedded in the urban environment or not – are still constantly innovating the way we move within a city, and therefore how we live in it. Smart public transport cards change the way users interact with public transit systems, and the data gathered by smart card processing drives the decision-making processes underlying its management. Smartphone signals \(^{73}\) permit to obtain geo-referenced data from urban dwellers, thus allowing e.g. granular, real-time traffic monitoring, faster incident response, or data-driven public transportation schedule adjustments. Energy provisioning is another domain that is directly shaped by the evolution of contemporary ICT: smart meters, for instance, allow for a much more granular knowledge of users’ habits and energy consumption patterns,\(^{74}\) thus ideally permitting a better resource allocation,\(^{75}\) and wiser, more informed policy and business decisions – some of the goals that smart cities are said to be pursuing.\(^{76}\)

The introduction and embedding of ICT in the urban environment – often in a way that makes running code a condition necessary for the built environment to function as intended\(^{77}\) – and the “datafication” that follows\(^{78}\) have a transversal impact\(^{79}\) on a multiplicity of aspects of both living in a city and governing and managing\(^{80}\) one. Urban governance touches upon education,

---

\(^{72}\) Kitchin and Dodge (n 10); Townsend (n 6).

\(^{73}\) See Steenbruggen, Tranos and Nijkamp (n 2); Townsend (n 15); Townsend (n 6); Eagle and Greene (n 16); Carlo Ratti and others, ‘Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis’ (2006) 33 Environment and Planning B: Planning and Design 727; Colin Harrison and others, ‘Foundations for Smarter Cities’ (2010) 54 IBM Journal of Research and Development 1. See also Bradshaw (n 15); Bernie Meyerson, ‘Smartphones May Enable Smart Cities’ Financial Times (4 March 2013).

\(^{74}\) Colette Cuipjers and Bert Jaap Koops, ‘Smart Metering and Privacy in Europe: Lessons from the Dutch Case’ in Serge Gutwirth and others (eds), European Data Protection: Coming of Age (Springer 2013).

\(^{75}\) “Cities rely on too many external resources and, as a matter of fact, they are (and probably will always be) consumers of resources. Promoting sustainability has been interpreted through the promotion of natural capital stocks. Other, more recent, interpretations of urban sustainability have promoted a more anthropocentric approach, according to which cities should respond to people’s needs through sustainable solutions for social and economic aspects” – Vito Albino, Umberto Berardi and Rosa Maria Dangelico, ‘Smart Cities: Definitions, Dimensions, Performance, and Initiatives’ (2015) 22 Journal of Urban Technology.

\(^{76}\) See e.g. Caragliu, Del Bo and Nijkamp (n 17); Townsend (n 6); Neirotti and others (n 8).

\(^{77}\) See Kitchin and Dodge (n 10).

\(^{78}\) See Kitchin, The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences (n 4).

\(^{79}\) “Smart city initiatives have a significant impact on various sides of a smart city (governance, people and communities, economy, natural environment, and built infrastructure). These are not only the aspects of outcomes made by smart city initiatives, but the components as contexts and conditions of localities also shape the characteristics of smart city initiatives” – Suha Alawadhi and others, ‘Building Understanding of Smart City Initiatives’, Electronic government (Springer 2012), 43.

\(^{80}\) “(W)ith the right software and dashboard in place, it’s possible to view data cross an entire physical infrastructure. In other words, an agency or organisation could clearly determine – based on structural data rather than opinions and politics – the real world risks and costs of fixing or ignoring a problem”: Samuel Greengard, The Internet of Things (MIT Press 2015) 66.
security and policing, emergency management, fostering economic and social development, utilities supply, works like streets and sewers management, waste disposal, transportation, and so on and so forth. The instrumentation of the built environment — cities turning ‘smart’ — has a horizontal effect, which interests directly or indirectly all the sectors of urban governance and the facets of city dwelling.

This chapter gives an account of the concept of ‘smart city’ — the embodiment of the progressive digitalization and instrumentation of the urban environment. In order to do so, it appears necessary to attempt at defining the concept of smart city, which has so far eluded a comprehensive, authoritative characterisation and a commonly accepted definition,\(^81\) briefly reviewing and synthesizing the classifications given by the relevant literature.\(^82\) In the rest of the dissertation, which is mainly dedicated to the relationship between open data and data protection, the smart city environment will be used as a contextual element to frame the setting of the analysis. In order to assess how smart cities can, in practice, dent individuals’ right to the protection of personal data, a deeper understanding of what a ‘smart city’ actually is needs therefore to be demarcated.\(^83\) At the same time, appreciating the drivers and consequences of the “datafication” and instrumentation of the built environment is crucial for understanding the role and value of open data for smart cities, and both its promises and its perils.

### 2.1 Making Sense of the Smart City

The concept of smart city represents, ultimately, the modern declination of the concept of city: a system of systems interacting with each other within a geographically localized area, forming an ever evolving and shifting entity in which an enormous multiplicity of actors interacts. Each city is different: urban development — both in its physical and in its social declination — is a diachronic process that starts with the initial settlement and continues throughout all the city’s lifespan, conditioned by the city’s own necessities and priorities and by a multiplicity of other factors.

Every city — and hence every smart city\(^84\) — is unique. At the same time, however, every city has some characteristics that are comparable to other cities’,\(^85\) some functions that are performed in a similar way. It is certainly possible to discuss cities as a general category, and to compare them, for instance, in function of their scale, but one must be mindful of the fact that each of

---


\(^{82}\) The execution of a literature review is meant both to avoid the rediscovery of what is already known and to stimulate a rigorous utilization of what already belongs to the state of the art: Jan Vom Brocke and others, ‘Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process’, *ECIS Proceedings* (2009).

\(^{83}\) “An understanding of the definition of a smart city is vital to be able to understand its scope and content”: ChuanTao Yin and others, ‘A Literature Survey on Smart Cities’ (2015) 58 Science China Information Sciences 4/18.

\(^{84}\) “Each deployment of ‘smart city’ technologies reflects local patterns of growth, urban governance models, and knowledge transfer networks” — Christopher Gaffney and Cerianne Robertson, ‘Smarter than Smart: Rio de Janeiro’s Flawed Emergence as a Smart City’ [2016] Journal of Urban Technology 1, 2.

them is the by-product of a number of characteristics that render it distinctive. Its environmental and geographical setting, for example, and its climate; its history, demographics, and social context; its written laws, social norms, and economy; its governance, and the division of powers and duties between local and national government, and also between the agencies operating within the city’s territory. And again, each city has its own actors and activities, its hard and soft infrastructure, its facilities and buildings, its own priorities and objectives.

The drive of this section is to examine the notion of smart city from a holistic and critical perspective. The section aims at enucleating its characteristics, and at reaching a working definition to be used for the purposes of this dissertation – identifying and tackling the conflictual points arising from the interaction between data protection and open data within the cities of the future. Several different terms have been adopted in the attempt to encompass, with a single buzzword, all the facets reflecting the progressive digitalization and networking of the urban environment. Smart city, intelligent city, sustainable city, digital city, real-time city, even Metropticon: the most commonly used definitions change according to the specific aspects of the digital instrumentation of the urban environment considered, and to the type and focus of the document in which the definition is included.

Albeit the complexity of the issue at stake makes a potential ‘one-size-fits-all’ definition unlikely to possibly be both comprehensive and synthetic at the same time, it still appears to be necessary to attempt at giving a holistic review of the building blocks of the smart city concept, within which to frame the clash between open data and data protection. The interaction between data protection and open data within the urban environment, on which the following chapters elaborate, can indeed potentially be starkly re-shaped by the transformation of traditional cities into ‘smart’ ones.

The purpose of the analysis is to help defining a notion of smart city that is both comprehensive enough not to exclude any major aspect of the progressive digitalization and networking of the urban environment and fit for this research’s conceptual framework, i.e. focused on the relationship between open data and data protection. A non-systematic (or narrative) literature review was therefore utilized in order to enucleate the main cores comprising a holistic notion of smart city. Specifically, other than the given definitions, the review took into consideration what the sources examined considered to be the building blocks of the smart

86 “(T)here is a clear problem conflating smart cities with a range of terms like cyber, digital, wired, knowledge cities etc., when in fact these various ideas themselves have somewhat different meanings”: Robert G Hollands, ‘Will the Real Smart City Please Stand up? Intelligent, Progressive or Entrepreneurial?’ (2008) 12 City.
87 E.g. Nam and Pardo (n 81); Dameri (n 81); Townsend (n 6).
88 See Nam and Pardo (n 81).
89 ITU-T Focus Group on Smart Sustainable Cities, ‘An Overview of Smart Sustainable Cities and the Role of Information and Communication Technologies’ (2014).
90 See Albino, Berardi and Dangelico (n 75); Cocchia (n 9).
91 See e.g. Townsend (n 15); Rob Kitchin, ‘The Real-Time City? Big Data and Smart Urbanism’ (2014) 79 GeoJournal 1.
92 Finch and Tene (n 21).
94 Bryman (n 68).
95 “Literature reviews are playing an increasingly important role in social scientists’ definition of knowledge” – Harris M Cooper, ‘Organizing Knowledge Syntheses: A Taxonomy of Literature Reviews’ (1988) 1 Knowledge in Society 104.
city environment – the elements that are making cities become ‘smart’. The purposes for which the traditional urban environment is shifting towards the smart city have also been considered by the review. In a nutshell, the review undertaken focused on terminology, definitions, and typologies, and on the objectives and purposes to be pursued by developing the smart city ecosystem, without yet applying any chronological of geographical sorting. The research within which the literature review takes place is carried out from a fundamental rights perspective, and adopts a teleological approach, thus espousing a defined position conditioned by the overarching aim of the thesis: conceptualising a satisfactory equilibrium between data protection rights and open data instances within the smart city context. The progressive commodification of individuals’ privacy and data protection rights is by no means a recent phenomenon, and is considered within this thesis’ conceptual framework, thus influencing the perspective from which the notion of smart city is investigated.

2.2 What is a Smart City?

Smart cities can be examined and portrayed focusing on a number of aspects, each of which depicts a different way in which the urban environment has been affected by the rampant digitalization and networking that characterized the past decades. Specifically, the research undertaken highlights that the existing literature on smart cities considers the topic under many different perspectives, and through the lenses of several different backgrounds and disciplines, which is to be expected considering how the intertwinement of ICT with the built environment interests several kinds of professional expertise. Additionally, the decision-making processes through which smart cities are run, appear to have a twin orientation, meaning that the flow of information underlying them can potentially move in two specular directions: top-down, or bottom-up. The literature reviewed considered the topic from two distinct points of view, i.e. a theoretical one, in which a high-level approach to the concept is dominating, and an eminently practical one, in which the intelligence of a city is largely equated to and dependent from the specific technologies concretely deployed in the built environment. Finally, from the review undertaken, it is possible to distinguish smart city models according to

---

96 The literature review followed a three-step approach adapted from the steps proposed by Vom Brocke and others (n 95) and from the ones by Bryman (n 68). After the initial development of a review strategy (step one) a number of databases were accessed and scanned for the relevant literature, which has then been classified, examined, and stored (step two); the third step was to draw conclusions from the analysis undertaken, using them to craft this dissertation’s understanding of the syntagm ‘smart city’, and thus a working definition of the concept. The literature coverage is representative, rather than exhaustive: it better fits the research in which the review is embedded, and avoids the repetition of notions too similar with each other to provide a meaningful contribution to our definition’s development. Regardless of the initial scope limitation, thanks to a number of literature reviews previously undertaken by other authors and organisations, more than 150 definitions were examined. The definitions considered, finally, were organised according to their conceptual proximity, i.e. clustered together according to the facet of the concept of smart city they portray.

97 Cooper (n 95).


their tangible implementation, i.e. concretely depending on how and at which point of the city’s development the ICT stack is integrated with the urban environment.

2.2.1 Perspectives on the Smart City

The networking and digitalization of the city’s environment has an impact that is horizontal, transversal to a number of understandings of what cities are. The literature examined appears to assume a multifaceted, threefold perspective on what a smart city actually is: smart cities can be framed (and have been framed) from an (1) technological perspective, (2) an organisational one, and (3) a human perspective, an anthropocentric standpoint that focuses mostly on the idea of “smart citizen”.

2.2.1.1 Technological Perspective on Smart Cities

As ICT is the minimum common denominator linking together all things ‘smart’ (e.g. smartphones, smart homes, smart TVs), and considering that data mining and analytics techniques are the very foundation of a new way of running and thinking about the urban environment, a number of definitions focus on the technological layer of the smart city. The embedment of an ICT stack and of networking capabilities in the built environment has been highlighted e.g. by Hollands, who notes that “(o)ne of the key elements which stands out in the smart (intelligent) city literature is the utilization of networked infrastructures to improve economic and political efficiency and enable social, cultural and urban development”; or by Forrester Research, according to which “(t)he use of Smart Computing technologies to make the critical infrastructure components and services of a city — which include city administration, education, healthcare, public safety, real estate, transportation, and utilities — more intelligent, interconnected, and efficient”.

What emerges from the review of the definitions enclosed in this category is that the notion of smart city is eminently technology-based, i.e. that technological development in ICT and data analytics, when intertwining with the built environment, constitute tout court the ‘smart’ in the ‘smart city’ syntagm. The technologically-oriented perspective transpires also from the literature that does not actually provide a definition or a general understanding of what a smart city is, rather offering, for instance, technical solutions aiming at solving single issues belonging to an individual smart city domain, or an overview of the state of the art in specific technology domains relevant to the smart city environment.

100 Nam and Pardo organise the smart city definitions considered in their inquiry into three categories, sorting them depending on a number of common elements; they, too, refer to a technological dimension, a human dimension, and an institutional dimension.
101 See Kitchin, ‘Data-Driven, Networked Urbanism’.
102 In particular, the corporations that turned out to be the most prominent players in the smart city market (e.g. IBM, Cisco, and Siemens) look at ICT as the cornerstone of their understanding of smart city: Albino, Berardi and Dangelico.
103 Hollands, p. 307.
104 Doug Washburn and others, ‘Helping CIOs Understand “Smart City” Initiatives’.
105 E.g.: “(t)he vision of “Smart Cities” is the urban center of the future, made safe, secure environmentally green, and efficient because all structures [...] are designed, constructed, and maintained making use of advanced, integrated materials, sensors, electronics, and networks which are interfaced with computerized systems comprised of databases, tracking, and decision-making algorithms” – B Bowerman and others, ‘The Vision of a Smart City’, 2nd International Life Extension Technology Workshop, Paris.
A large number of technologies and of broader technological trends have been identified as core to the notion of smart city. Ubiquitous computing – often also referred to as pervasive computing, ambient intelligence,107 or “everyware” 108 – strives to enable computation everywhere and through any device. Broadband networking and cloud computing109 remove the constraints which information was bound to before – low bandwidth and local storage and processing. Big data110 technologies allow to process high-dimensional, complex and constantly updating data sets, which are increasingly often released as open data – a policy that mandates public administrations to release their data in an accessible manner for all to use for every purpose. Large arrays of distributed and networked sensors and actuators embedded in several objects and devices – the Internet of Things (IoT) – allow to gather huge and varied amounts of data, often in real time, with increasing granularity and detail. GISs (Geographic Information Systems) and BIM (Building Information Modelling) tie the spatialities of the smart city to its informational components, allowing its digital representation and modelling. E-Government services provide a new interface between the city’s administration and its citizens, linking them through ICT infrastructure and services.111

Most of the smart city definitions and notions that were examined during the research undertaken have in ICT a common element, sometimes as their very core, some other times merely as a component – necessary but not in itself sufficient – to be present when qualifying a city as ‘smart’. Indeed, as it has been highlighted,112 integrating and mediating urban systems with ICT is hardly a new idea per se, nor is it unique to the smart city environment: what characterizes the smart city concept is a holistic approach to urban management, governance, and dwelling, given by a wide array of technologies and processes bound together by an overarching set of goals.

2.2.1.2 Organisational perspective on smart cities: governance and management

Part of the literature examined highlights how cities becoming smart also means – either incidentally or principally – a shift in the organizational and decisional practices on which urban governance, management and development are based.113 Urban governance and planning are bound to become evidence-based, data-driven,114 as smart as cities are: “governing a smart city is about crafting new forms of human collaboration through the use of information and communication technologies [...] technology by itself will not make a city smarter: building a

110 Mayer-Schönberger and Cukier (n 13).
111 See ISO/IEC JTC1 (n 85).
113 See e.g. Alawadhi and others (n 79); Taewoo Nam and Theresa A Pardo, ‘Smart City as Urban Innovation: Focusing on Management, Policy, and Context’, Proceedings of the 5th international conference on theory and practice of electronic governance (ACM 2011); Nam and Pardo (n 81).
114 Kitchin, ‘Data-Driven, Networked Urbanism’ (n 3).
smart city requires a political understanding of technology, a process approach to manage the emerging smart city and a focus on both economic gains and other public values”.  

The smart city, under this perspective, is more than the sum of the technologies employed within its boundaries: it is a shift towards different governance frameworks, a new approach on urban management, based on the data gathered by the array of sensors the city is instrumented with, and then further processed by its computing infrastructure.

In the large majority of the instances examined, even when the focus lies on the governance and organizational aspects of the smart city environments, ICT and analytics still occupy a prominent role within the definition of smart city: “(u)rbang big data, city operating systems, urban informatics, and urban science analytics provide the basis for a new logic of urban control and governance -- data-driven, networked urbanism -- that enables real-time monitoring and steering of urban systems and the creation of what has widely been termed smart cities. The notion of a smart city can be traced back to experiments with urban cybernetics in the 1970s, the development of new forms of city managerialism and urban entrepreneurship, including smart growth and new urbanism, in the 1980s and 90s, and the fusing of ICT and urban infrastructure and development of initial forms of networked urbanism from the late 1980s onwards”.

The fact that a comprehensive smart city definition needs to take into account urban governance and management is underlined by a relevant chunk of the literature reviewed, which emphasises how smart cities are governed by a defined group of subjects, clearly identifiable and able to set rules and policies for urban management and development.

2.2.1.3 HUMAN PERSPECTIVE ON SMART CITIES: HUMAN CAPITAL
The organisational perspective, which highlights how smart cities can also be conceptualized in terms of a revolution in urban planning and management, is closely linked to what could be defined as a human perspective on the smart city, an anthropocentric approach that emphasizes how city dwellers are both the main beneficiaries and the main agents for and through which cities are turning smart, and how human capital is one of the main drivers behind the shift towards ‘smarter’ city.

City dwellers are arguably bound to be what cities revolve around, both in their capacity as individual human beings, and as collectively belonging to those social formations in which individualities aggregate. According to some of the definitions belonging to this standpoint, smart citizens are a major driver – even the main one – pushing cities’ collective intelligence

116 Kitchin, ‘Data-Driven, Networked Urbanism’ (n 3).
117 See Dameri (n 81) 2549.
118 “(W)hat defines the smart city is not the infrastructures or networks it offers, but the ways in which its citizens interact with these systems as well as each other”: Nils Walravens, Jonas Breuer and Pieter Ballon, ‘Open Data as a Catalyst for the Smart City as a Local Innovation Platform’ [2014] Communications & Strategies 20.
119 See e.g. Caragliu, Del Bo and Nijkamp (n 17).
120 See e.g. Alawadhi and others (n 79). See also Albino, Berardi and Dangelico (n 75) 9: ‘[p]eople are the protagonists of a smart city, who shape it through continuous interactions. For this reason, other terms have often been associated with the concept of the smart city. […] The social infrastructure, such as intellectual and social capital, is an indispensable endowment to smart cities as it allows “connecting people and creating relationships”’. 
121 See e.g. Townsend (n 6).
forward. As it has been noted, “the issues for the creative city of the future will focus upon its ‘soft infrastructure’, including such things as knowledge networks, voluntary organizations, safe crime-free environments and a lively after dark entertainment economy [...] This more ‘humanist’ emphasis ties in with other related discourses of smart communities, including the importance of social leaning, education and social capital for developing the smart city”. Even corporate literature considers humans, both as individual citizens and as belonging to a wide array of social formations and networks, as a system (rectius, perhaps, as a subsystem) within the ones that compose a smart city.

In a nutshell, smart cities appear to be understood by the literature reviewed according to three different (yet closely connected) perspectives, each of which assigns a different weight to the factors upon which smart cities are built. The technological perspective reigns sovereign, its prominence hardly questioned by the literature reviewed. Moreover, the digitalization and networking of the urban environment has a direct impact on those urban management, governance and organisational practices through which cities are run, thus allowing to highlight how a city’s intelligence is not just a technological issue, but also an organisational one. Finally, cities are inhabited by humans, run by people, and largely shaped according to how individuals interact within them; accordingly, the examination performed and all the other literature reviews considered highlighted the presence of an anthropocentric perspective on the concept of smart city, focused on the smart citizen – informed, creative, inclusive and included people – and on their role in the growth of the cities of the future.

Despite the centrality of ICT and of technological innovation in the built environment in general, the notion of smart city does not merely equate with the application of modern technologies to the urban setting, but considers people and communities as well, and the ways in which they organize and are managed too. Despite the fuzziness resulting from the heterogeneity of the aspects of the smart city context, six dimensions appear to be shared amongst most smart city definitions or benchmarking models: people, government, economy, mobility, environment and living; see Leonidas G Anthopoulos, Marijn Janssen and Vishanth Weerakkody, ‘Comparing Smart Cities with Different Modeling Approaches’, Proceedings of the 24th International Conference on World Wide Web Companion (International World Wide Web Conferences Steering Committee 2015).

122 Despite the fuzziness resulting from the heterogeneity of the aspects of the smart city context, six dimensions appear to be shared amongst most smart city definitions or benchmarking models: people, government, economy, mobility, environment and living; see Leonidas G Anthopoulos, Marijn Janssen and Vishanth Weerakkody, ‘Comparing Smart Cities with Different Modeling Approaches’, Proceedings of the 24th International Conference on World Wide Web Companion (International World Wide Web Conferences Steering Committee 2015).

123 Hollands (n 86) 309.


125 “ICT is central to the operation of the future city”: Batty and others (n 10).

126 Even if the smart city concept deals with innovation in the urban space in general, and innovation does not necessarily have to be ICT based: see Anthopoulos, Janssen and Weerakkody (n 122).

127 “(T)he smart city concept is no longer limited to the diffusion of ICT, but it looks at people and community needs”: Albino, Berardi and Dangelico (n 75) 5.


129 See Albino, Berardi and Dangelico (n 75) 4.

130 Albino, Berardi and Dangelico (n 75) 5.
2.2.2 The Smart City as a Regulatory Instrument

It is by now pacifically recognized how technology functions as a regulatory instrument. Its physical dimension (architecture),\(^{131}\) its digital, immaterial counterpart (code),\(^ {132}\) and their merger (code/space),\(^{133}\) have the potential to shape and regulate human behaviour as much as the law or social norms have – in many instances, even more. Let us exemplify: to curb cars’ speed in a residential neighbourhood, a local administration could rely solely on regulation by law (e.g. setting a very low speed limit and a high speeding fine) or on architectural design, e.g. by placing speed bumps, speed traps, or bottlenecks. To keep a malicious intruder out of a computer network, one could rely on the set of countervailing norms and sanctions through which the law criminalizes unauthorised access, or deploy an intrusion detection system as well. In both cases technology – architecture and code – can be used to substitute or integrate enforcement by other regulatory instruments, such as the law, social norms, and the market.\(^{134}\)

Architecture, in particular, has been shown to perform a regulatory function by expressing and imposing cultural or symbolic meanings; by directly affecting how people interact; and by being biased – either positively or negatively – towards certain social groups, values, or practices.\(^ {135}\) Architectural constraints can be employed to communicate and convey values and by either promoting or being sensitive to particular biases. They can also constrain or direct human behaviour, influencing social interaction and ordering; define personal space and territoriality, manipulating an area’s spatialities; dominate, discipline, and reform individuals,\(^{136}\) as exemplified by Jeremy Bentham’s panopticon.\(^ {137}\)

At the same time, technologies – artefacts – are inherently political:\(^ {138}\) they embody a set of pre-existing values deriving from the choices of the ones that engineered them. The fact that they can determine, change, and coerce human conduct makes it so that, if their scope is sufficiently wide, their regulatory capacity shapes both individual and collective behaviour according to the underlying values transferred by who designed or deployed those technologies. The instrumentation of the built environment – the “datafication” of urbanities, the smart city – relies on technologies of such a scale and regulatory capacity. Again, let us exemplify. To promote sustainable growth and efficiency, a local public administration might decide to instrument rubbish bins and rationalize waste collection. It might decide, for instance, to use sensors to detect when the bin is at capacity, hence alerting waste collection operators only when necessary, saving some expenditures to the city. It might also decide, however, to instrument those bins with access control mechanisms so that only households that have paid waste disposal taxes have access to it, or with a sensor system designed to


\(^{132}\) Lessig, Code and Other Laws of Cyberspace (n 67); Ronald E Leenes and Bert-Jaap Koops, “Code” and Privacy - Or How Technology Is Slowly Eroding Privacy’ in Egbert Dommering and Lodewijk Asscher (eds), Essays on the Normative Role of Information Technology (TMC Asser Press 2005); Katyal (n 131).

\(^{133}\) Kitchin and Dodge (n 10).

\(^{134}\) Lessig, Code and Other Laws of Cyberspace (n 67).

\(^{135}\) Shah and Kesan (n 131).

\(^{136}\) Shah and Kesan (n 131).

\(^{137}\) See Michel Foucault, Discipline and Punish: The Birth of the Prison (Vintage 1977).

identify (and then fine) who violates recycling norms.\textsuperscript{139} The artefacts that instrument the individual systems (e.g. waste disposal, energy provisioning) that conjoined form that system of systems we define as city have regulatory capacity, are a political issue,\textsuperscript{140} and embody an underlying set of values. The realist epistemology through which the smart city is portrayed as a mere stack of neutral technologies is a misleading narrative: the instrumentation of the built environment actively translates certain values into reality,\textsuperscript{141} and regulates human behaviour.

There appear to be two main conceptions,\textsuperscript{142}, regarding the regulatory capacity and the “direction of travel” of the information processed within a smart city.\textsuperscript{143} The first one can be sketched as a top-down standpoint, echoing the idea of an integrated command-and-control operation centre that makes sense of the data flows resulting from the extensive array of networked sensors deployed within the smart city environment, a ‘control room’\textsuperscript{144} for the whole city. The second orientation the data flows underlying smart cities could have is a bottom-up one,\textsuperscript{145} driven both by the availability of a myriad of ubiquitous computing devices – e.g. smartphones, tablets, laptops, etc. – working as granular, mobile, inexpensive hand-held sensors, on one hand, and by the instances aiming at fostering inclusions and participatory governance that underlie the very concept of smart city,\textsuperscript{146} on the other.

\subsection{Top-down Regulatory Capacity}

The idea of a control room\textsuperscript{147} for the built environment is furthered by a number of sources belonging to the corporate world,\textsuperscript{148} which aim at presenting and – eventually – at selling and implementing comprehensive, integrated solutions targeted towards modern urban management needs. Rio de Janeiro, whose municipality was one of the first to implement IBM’s Intelligent Operations Centre,\textsuperscript{149} appears to be a prime example of a top-down viewpoint

\footnotesize
\begin{itemize}
\item\textsuperscript{139} The example is inspired by Liesbet van Zoonen, ‘Privacy Concerns in Smart Cities’ (2016) 33 Government Information Quarterly.
\item\textsuperscript{140} See Jathan Sadowski and Frank A Pasquale, ‘The Spectrum of Control: A Social Theory of the Smart City’ (2015) 20 First Monday. See also Jeroen van den Hoven, ‘Architecture and Value-Sensitive Design’ in Claudia Basta and Stefano Moroni (eds), Ethics, design and planning of the built environment (Springer 2013).
\item\textsuperscript{141} “Information technology has become a constitutive technology and partly constitutes the things to which it is applied. It shapes our discourses, practices and institutions and experiences in important ways”: Noëmi Manders-Hufts and Jeroen van den Hoven, ‘The Need for a Value-Sensitive Design of Communication Infrastructures’ in Paul Sollie and Marcus Düwell (eds), Evaluating New Technologies (Springer 2009) 68.
\item\textsuperscript{142} not necessarily mutually exclusive, see: “Change seldom arises from purely top-down or bottom-up systems and processes, and pitching each paradigm in opposition to the other simply refies [sic] their shortcomings”: Mark Shepard and Antonina Simeti, ‘What’s So Smart About the Smart Citizen’ in Drew Emment and Anthony Townsend (eds), Smart Citizens, vol 4 (Future Everything Publications 2013) 17.
\item\textsuperscript{143} See Walravens, Breuer and Ballon (n 14).
\item\textsuperscript{144} E.g. see Pam Nesbitt, ‘IBM Intelligent Operations Center for Smarter Cities’ (2012); Shepard and Simeti (n 142).
\item\textsuperscript{145} See Drew Hemment, Anthony Townsend et al., Smart Citizens (Drew Hemment and Anthony Townsend eds, FutureEverything Publications 2013).
\item\textsuperscript{146} See inter alia Caraglua, Del Bo and Nijikamp (“sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance”). See also Dameri, who talks about “benefits for citizens in terms of well being, inclusion and participation, environmental quality, intelligent development”.
\item\textsuperscript{147} See Walravens, Breuer and Ballon (n 14); Townsend (n 6); Nesbitt (n 144).
\item\textsuperscript{148} E.g. Harrison and others (n 73).
\item\textsuperscript{149} Which reportedly sports an 80 square meter high-definition video wall that displays live information from 560 cameras, a weather forecasting system, and a map capable of analysing 60 different layers of data coming from the sensors embedded within the city’s environment; more than 20 agencies were integrated within the
\end{itemize}
of the smart city ecosystem, where ICT is embedded within the built environment to optimize its functioning, overseeing urban activities, gathering information, and interacting – potentially autonomously – with the city’s infrastructure. In Rio, IBM’s smart city solutions have been implemented in two smart city centres: the Centro Integrado de Comando e Controle (CICC-RJ) and the Centro das Operações do Rio (COR). The CICC-RJ is Rio’s security operations centre, which concentrates the city’s security apparatus, and is staffed with members of Rio de Janeiro’s military police. Its functions are security-oriented, and include running Rio’s emergency services, traffic and public transportation monitoring, and acting as the city’s security planning and operations hub. As opposed to the CICC-RJ’s security function, the COR has a civilian nature: it monitors traffic flows and weather patterns and coordinates Rio’s civilian agencies, systems, and services. Ultimately, this top-down approach to the smart city environment seems to be an eminently infrastructural point of view, closely linked to the solutions marketed and sold by major technology providers operating in the smart city business space. In this regard, it appears appropriate to discuss the tendency, which started in the mid 1990s, to benchmark a city’s performance through a multiplicity of urban indicators and then to report and represent them visually through dashboards.

This kind of measurement, benchmarking and visualization of the urban environment – a distinctive characteristic of the modern smart city as portrayed by part of the literature reviewed – is expressive of a top-down approach to its contextualization. As it has been noted, on one hand, urban indicators, benchmarking, and dashboards enable or at the very least facilitate data-driven governance and evidence-based decision-making, as opposed to opaque political processes. On the other hand, however, the representation of the urban environment through artificial metrics – as any technological advancement of sufficient reach and scope – is to be considered within its broader social and political context: technology is seldom neutral.

Cities are too complex and multifaceted to be entirely represented as a collection of data points; any technological system monitoring and measuring a city’s performance and indicators is not merely translating that city’s characteristics into measurable and comparable information, but is actively contributing to its framing and future development. Kitchin et al. underline that urban indicators, benchmarks, and dashboards are not just toolkits, but data assemblages – socio-technical systems “composed of many apparatuses and elements that are thoroughly entwined”, and that rather than offering a neutral portrait of a city’s reality are actively producing it. Indicators, benchmarks, and dashboards reflect a top-down approach to the smart city in that both their architectural design and their interpretation are expression of the choices made by the local government and by who developed them.

Intelligent operations centre, decreasing average response time by 30%: Jenny Soffel, ‘Rio’s “big Brother” Control Room Watches over the City’ CNN (29 August 2013).

150 Walravens, Breuer and Ballon (n 14).
151 See Gaffney and Robertson (n 84).
152 Walravens, Breuer and Ballon (n 14).
154 Kitchin, Lauriault and McArdle (n 153) 15 ss.
155 Kitchin, Lauriault and McArdle (n 153).
156 Kitchin, Lauriault and McArdle (n 153) 17; Kitchin, The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences (n 4) 24 ss.
157 Kitchin, Lauriault and McArdle (n 153) 17.
Let us consider, as an example, the placement of an interconnected array of sensors that monitors the air quality of a conurbation as part of a particular smart city initiative. The deployment would have to be conditioned to the decisions of the administration with regard to the kind of sensor to use and where to place them: what will the sensors be designed to detect? Will they be deployed only in the city centre, or in the low-income suburbs too? The output would then have to be analysed through algorithms and formulas whose design cannot be neutral, and is the by-product of a series of decisions taken by their creators; it would also have to be visualized through software that shares the very same characteristics, and whose functioning reflects the choices (and the underlying values) of its creator. Decisions would then be made in consideration of the result of what it has been decided to benchmark, measure, and represent. Local government bodies could for instance decide to allow or deny the building permissions for an industrial facility in consideration of the sensors’ readings, of their visualization, and of the urban indicators and benchmarks chosen. Indicators and benchmarks, sensors and real-time dashboards, embedded in their context of operation, are part of a socio-technical system whose rules are ultimately set by the urban government and by its technology providers. In this sense, the ICT that enables the smart city’s functioning can be seen as part of a top-down approach to the construction of the smart city. The instrumentation of the built environment, viewed this way, is expression of a top-down way of regulating through technology: code and architecture come together to shape human behaviour according to a predefined set of decisions and underlying values on which individual city dwellers have often very little to say.

There is, arguably, the tendency to lean on a narrative that frames the smart city’s epistemology as “realist”, pushing the idea of the neutrality and objectiveness of the information gathered and processed by the smart city’s ICT infrastructure. Such a narrative is dangerous: it allows to falsely portray a top-down, command-and-control approach to the smart city concept as completely neutral and objective, thus masking the inherently political nature of smart city initiatives. As it has been rightly put, “(i)f the “smartest city” is the most liveable and just city, then we must ask how new technologies are working to achieve that goal. We must also ask what the motivations are for incorporating the evaluation frameworks and technology assessments of vendor as public policy metrics”.

2.2.2.2 BOTTOM-UP REGULATORY CAPACITY

Conversely, smart cities (and potentially the decisional processes underlying them) can also be envisioned as running on a constant data stream deriving by the advances in ubiquitous and mobile computing, coming directly from the smart devices carried by city dwellers, and thus

---

160 See Lessig, Code and Other Laws of Cyberspace (n 67); Lessig, Code Version 2.0 (n 67); Shah and Kesan (n 131); Kitchin and Dodge (n 10).
161 Gaffney and Robertson (n 84) 6.
162 Gaffney and Robertson (n 84) 6.
163 Gaffney and Robertson (n 84) 13.
– arguably – straightforwardly embodying the participative and inclusive instances which are central to many definitions of the smart city notion. Mobile smartphones, in particular, offer a great opportunity to understand e.g. mobility patterns and flows, often nearly in real-time, through aggregated data which comes directly by the subjects acting and dwelling within the urban environment.

In particular, the smart city infrastructure’s features can be synthetically summarized according to the following characteristics:

1. Real world awareness, deriving from the automated real-time sensing capabilities provided by an extensive network of interconnected sensors;
2. Knowledge-based systems (and the knowledge engineering and management capabilities necessary to build and implement them) to understand, analyse, and eventually act upon, the data gathered by the extensive network of sensors sub point 1;
3. A holistic view of the data gathered, purposed to interlink information pertaining to different domains in order to gain valuable insight to be used for governance and decision-making purposes.

Smartphones perfectly fit the characteristics to be sought in the smart city’s sensors’ infrastructure: mobile, pervasive, widespread, the sensors with which they are equipped boost the environment’s sensing capabilities, thus providing the real-world awareness smart cities require ex point 1 above. The cloud infrastructure on which smartphones rely, moreover, couples their mobile sensing capabilities with a backend that scales seamlessly and allows the offloading and processing of a massive amount of information, thus satisfying the requirement sub 2 above. Finally, the backend mentioned allows for the performance of advanced analytics, and thus for the ability to link the by-product of the data collected and analysed across different domains, gathering valuable insights on causations and correlations that would not have been apparent otherwise. It is indeed not the infrastructures and the architecture of the city’s ICT stack that make it ‘smart’: intelligence largely derives from the application layer.

The technical characteristics of modern smartphones (and of the apps for which they serve as platforms) and their widespread adoption and exceptional mobility and ubiquity make them therefore prime candidates to be exploited as building blocks of the smart city architecture; however, there is an additional important reason why smartphones are a perfect fit for the smart city infrastructure’s ideal characteristics: the close connection with the person they belong to. People, in their capacity of urban dwellers and as human capital, are a central, fundamental aspect of the smart city environment – an environment whose purpose is, amongst others, to foster inclusion and participatory governance. As mentioned, the smart city is eminently data-driven, meaning that decision-making processes are either based upon data and mediated by humans or directly taken by algorithms on the basis of the

---

165 Balakrishna (n 164). 225-226.
166 See, amongst corporate sources, GSMA, ‘Keys to the Smart City – How Mobile Operators Are Playing a Crucial Role in the Development of Smart Cities’ (2016).
167 See Eagle and Greene (n 16).
168 See Caragliu, Del Bo and Nijkamp (n 17).
170 For instance, traffic management based upon aggregated location data.
available information. Mobile computing devices are a conglomerate of sensors carried almost everywhere by a consistent chunk of the world’s urban population; they enable the gathering of information to be provided directly (albeit mediated by the device itself) by the subjects to which the benefits to be achieved by the smart city are directed, thus arguably including them indirectly – rectius, their data output – in the city’s governance processes. Through smartphones, the ‘datafied’ behaviour of the city’s inhabitants fuels the information pool on the basis of which both mediated and automated data-driven decision-making processes are carried out: smartphones act as a prime tool through which to foster the bottom-up dimension of the smart city environment. The ways through which the data they gather is processed and the processing’s very purposes, of course, ultimately condition how and whether the information collected is actually used to foster the inclusiveness and participatory governance of the urban environment of reference. However, the fact that smartphones collect the raw data that can be used to fulfill smart city objectives directly from city dwellers is arguably a fundamental element to be considered when framing the bottom-up dimension of the notion smart city.

The narrative pushed by major smart city technology vendors has been undergoing a polar shift in its direction of travel, flowing from a storyline that used to favour a top-down, command-and-control narrative, to an account that highlights a bottom-up, decentralized approach. It remains to the interpreter to evaluate which part of this narrative derives from the actual technological and social evolution of the smart city concept and of its implementation, or whether it is but vendors’ rhetoric, aiming at convincing decision makers regarding the validity of the integrated smart city solutions offered by large technology providers. It has indeed been pointed out how, as the next section will contribute in clarifying, the major corporate players did not, in fact, stumble upon an already existing market, with its needs and related opportunities. Rather, they actively created and shaped such a market, pushing the idea of ‘smartness’ as a desirable (and thus marketable) attribute of the future urban ecosystem, to be pushed and sold to investors and decision-makers operating in that space.

2.2.3 VISIONS OF A SMARTER CITY

The narrative about smart cities appears to be split between several different visions, distinct approaches whose adoption starkly impacts how smart cities are perceived by the various stakeholders of reference. On one hand, academia – urban studies, and humanities and social sciences in particular – and civil society seem to be able to often maintain a more critical approach towards the concept of smart city, often underlining its fuzziness, its potential pitfalls, and the fact that it is the concept’s real-world implementation that actually matters, rather than its mere framing. On the other hand, smart city initiatives are shaped by other key stakeholders that have been very active in trying to mould the narrative surrounding the notion

---

171 For instance, a smart lamppost automatically detecting the ambience’s lighting and regulating itself consequentially.
172 Kitchin, ‘Making Sense of Smart Cities: Addressing Present Shortcomings’ (n 93).
173 See Ola Söderström, Till Paaasche and Francisco Klauser, ‘Smart Cities as Corporate Storytelling’ (2014) 18 City.
of smart city. Technology vendors\textsuperscript{175} and consultancies,\textsuperscript{176} for instance, have a prominent role in steering the discourse around what smart cities are, while having all sort of vested interests in the instrumentation of the urban environment. Indeed, the smart city concept seems to have developed, to a large extent, thanks to corporate strategies and marketing tactics,\textsuperscript{177} which have had their own shaping effect on the concept of smart city as commonly understood nowadays.

Surely, an unambiguous difference between the approach assumed by business and consultancy research on a side and academia and civil society on the other is common across numerous disciplines and sectors – the aims of the two being radically different by definition. However, in the specific smart city context, the difference appears to be particularly marked. Academic critical studies question the very existence of the smart city concept as a separate notion from the city as traditionally understood, and dissect the whole narrative underlying it, often assuming – or attempting to assume – a holistic approach. The technical literature considered seemed mostly concerned with the development of particular technologies to be applied to specific domains, each of which directed at improving one or more sectors of urban living, governance and management. The corporate literature,\textsuperscript{178} on the other hand, seems to assume smart cities to be either already here or right about to, largely equating them with the deployment of ICT solutions, integrated within the urban environment, and interlinked by some sort of hub, control room, or operation centre.\textsuperscript{179} Indeed, it is to be expected by large technology companies and consultancy firms to try and market the smart city ideal to the subjects involved in urban governance as something that can be reached by simply conveying capital towards the deployment of ICT systems and towards the services surrounding it.

Think tanks, associations and alliances, and standard-setting organisations also play a role in shaping the concept of smart city, and hence the modern urban environment, to some extent. These kinds of stakeholder have been producing a sort of ‘grey literature’\textsuperscript{180} of reports, concept papers, and the likes, which seem to have the main purpose of advancing the specific agenda that the think tank or association is meant to push forth. For the open data community, for instance, smart cities are about data interoperability, sharing, and re-use, while for environment-focused organisations the prime concern is for smart cities to be green cities, and so forth.

Finally, cities themselves and the various administrative units by which they are composed are obviously very active in shaping the concept of smart city, both by producing their own reports and white papers and by way of public procurement.\textsuperscript{181} Either by aspirational statements of

\textsuperscript{176} See e.g. McKinsey’s report by Jonathan Woetzel and others, ‘Smart Cities: Digital Solutions for a More Livable Future’ (2018).
\textsuperscript{177} IBM’s “Smarter Cities” initiative, or Microsoft’s CityNext one, being prime examples.
\textsuperscript{178} E.g. Dirks and Keeling (n 124); Dirks, Gurdgiev and Keeling (n 124); Harrison and Donnelly (n 124); Harrison and others (n 73). More recent initiatives (e.g. the Sidewalk Toronto consortium) are seemingly diverging from the ‘technocratic’ path, as there seems to be a rising tendency to proper stakeholder engagement, which often highlights ‘softer’ side of urban governance.
\textsuperscript{179} See Nesbitt (n 144); Soffel (n 149).
\textsuperscript{180} Tracey Lauriault and others, ‘Open Smart Cities in Canada: Environmental Scan and Case Studies’ (2018) 5.
\textsuperscript{181} See e.g. the request for proposal for Toronto’s eastern waterfront issued by Waterfront Toronto, ‘Request for Proposals: Innovation and Funding Partner for the Quayside Development Opportunity RFP No.: 2017-13’ (2017).
purpose, branding and marketing, or public procurement and requests for proposals, many metropolitan cities have been frontrunning the smart city revolution. The role of cities in shaping themselves is so obvious it seems almost a tautology. However, as the smart city narrative entrenches itself in the public discourse about the future of urbanities, it almost looks as if there were a single ‘smart city’ model to be followed or set of requirements to be achieved\textsuperscript{182} for a city to attain ‘smartness’, while in reality each city has its own particular goals, challenges, and constraints, and hence its vision of what ‘smart’ means.

So where is the smart citizen? The dominant smart city narratives contain plenty of aspirational statements on the role of ‘smart citizens’ in the smart city, and the proliferation and increased accessibility of sensors and computing technologies has undeniably led to advances in citizen sensing and citizen science\textsuperscript{183} that has bolstered the role of individual inhabitants in urban research and participation. Some research suggests that the concept of ‘smart citizen’ does not really match the level of awareness and engagement that the common citizen actually has,\textsuperscript{184} while other authors argue that – at least in some specific cases – citizens are actually becoming active decision-makers, rather than mere data providers.\textsuperscript{185} Recent large-scale smart city initiatives, such as the Sidewalk Toronto project,\textsuperscript{186} or Barcelona’s smart city initiative, have been increasingly engaging in different forms of dialogue with the citizens of the areas involved. All in all, it seems that the contribution citizens can have, directly and as individuals, to the development of the ‘smart city’, both as a social construct and in its concrete implementation in real world scenarios, is still largely happening through the groups in which individual citizens organise.

2.2.4 IMPLEMENTATION OF THE SMART CITY

Cities hardly fit a one-size-fits-all category, and their smartness is no exception: each smart city – assuming there actually is such a thing – is a distinct entity, its characteristics potentially different from other cities, focusing on a different set of problems to be solved or mitigated through the deployment of ICT solutions within the urban environment. One of the distinctions to be made when framing specific smart city examples – and thus to be considered when attempting at defining what a smart city is – relates to how the ICT solutions are embedded in

\textsuperscript{182} One of the postulates underlying the research questions of this thesis, and of the grant agreement that funds the project on which the research questions are based, for instance, seemed to assumed that there were a set information required by cities to become ‘smart’, and that the EU data protection regime could have hampered the release of such information as open data. As it turned out, there are no specific requirements to be attained or categories of data to be processed – smartness is contextual to the challenges each city faces, and to the solutions it chooses to tackle them.


\textsuperscript{184} Vanessa Thomas and others, ‘Where’s Wally? In Search of Citizen Perspectives on the Smart City’ (2016) 8 Sustainability 207.


\textsuperscript{186} See https://sidewalktoronto.ca/documents/ for a number of documents reporting on the outcome of roundtables, panels, and other kinds of public engagement processes.
the conurbation considered. It has been indeed highlighted how not every smart city is implemented the same way: the urban environment can be retrofitted and renovated to fit the ‘smart’ paradigm *ex post*; it can undergo a ‘shock therapy’, where ICT is rapidly and forcibly integrated into an existing ‘dumb’ landscape; it can be built *ex novo* as a smart city, as e.g. South Korea’s New Songdo has. It also appears appropriate to add how there is the chance that a smart city could be merely portrayed as smart, as a result of a marketing stunt or publicity campaign, while remaining a traditional kind of conurbation: the smart city concept largely derives from corporate instances, made prone to mystification by the fuzziness of its (many) definitions and qualifications.

The first kind of implementation – retrofitting existing cities to make them smart – is the most common one: converting their infrastructure, retrofitting and renovating the urban environment as traditionally considered in order to make it permeated with ICT, networked, and data-driven. This kind of implementation is bound to be a slow and extremely resource-intensive diachronic process, tied to specific necessities to be solved by what derives from the deployment of urban ICT solutions: this kind of conurbation gets ‘smart’ gradually, and only when there is an underlying reason justifying the budgetary allocation. This kind of implementation appears to be the most commonplace: cities, as they always did, face a specific (set of) problematic(s) whose answer is sought in a particular (set of) information and communication technology solution(s).

Closely connected to the first kind of smart city deployed is the one Sadowski and Pasquale refer to as “smart shock”, “wherein a city undergoes a quick, large-scale integration of ‘smart’ ideals, technologies, and policies into an existing landscape”. It might be difficult to distinguish between the first kind of implementation mentioned above and “smart shock” deployments: what matters in this respect is the pace of the transformation in consideration, and the way the shift from ‘dumb’ to ‘smart’ is integrated into the existing infrastructures and policies through which conurbations are run. An example could be given by the plans that the consortium formed between Waterfront Toronto and Alphabet’s Sidewalk Labs has for Toronto’s Eastern Waterfront, particularly the Quayside area. The project involves a 12-acre part of the city of Toronto, currently underdeveloped but still prime real estate, for which the aim is to “create and fund a globally-significant community that will showcase advanced technologies, building materials, sustainable practices and innovative business models that demonstrate pragmatic solutions toward climate positive urban development”. The consortium aims at rebuilding that part of the city in a radical manner, combining “forward-thinking urban design and new digital technology to create people-centred neighbourhoods that achieve precedent-setting levels of sustainability, affordability, mobility, and economic opportunity”. Cities are man-made constructs, and as such can be built completely *ex novo*. As mentioned, New Songdo in South Korea exemplifies this kind of smart city implementation, which consists

187 Sadowski and Pasquale (n 140).
188 See Shelton, Zook and Wiig (n 169).
189 Sadowski and Pasquale, 3/22.
190 See https://sidewalktoronto.ca/.
191 Waterfront Toronto (n 181) 6.
192 https://sidewalktoronto.ca/.
193 Sadowski and Pasquale (n 140); Manu Fernandez, ‘Smart Cities of the Future?’ in Drew Hemment and Anthony Townsend (eds), *Smart citizens* (FutureEverything Publications 2013).
of a newly built city in which ICT is integrated within the built environment from the city’s very beginning. The growing amount of population moving from the countryside to conurbations made it so that cities kept being built from the ground up even in modern times, as exemplified e.g. by Italy’s Latina, built during fascist times after the drainage and reclamation of the Agro Pontino wetlands, or by Brazil’s federal capital, Brasilia, built in less than four years (1956-1960) from the ground up. Even nowadays entire urban environments are built from scratchs, frequently due to top-down governance decisions, as exemplified, for instance, by the proliferation of China’s ‘ghost towns’. 

Finally, a smart city’s implementation can be merely nominal – a marketing campaign or publicity stunt aiming at steering the public opinion towards the perception of a given city. Smart cities have been cautiously framed by the available literature as a potential exercise in meagre corporate storytelling, an ‘innovation’ often largely constituted by a shift in political language and narratives, rather than by the conceptualization and deployment of an actually disruptive set of innovative practices and technologies purposed at revolutionizing how we manage and dwell within the urban environment. Indeed, the literature reviewed highlights how the smartness of a city does not come from the simple deployment of ICT solutions, however interconnected. It also involves a shift towards data-driven agency from an organisational perspective, and an impact on citizens’ practices too. Any city marketing itself as smart through the mere employment of disconnected ICT solutions, and thus lacking a holistic approach, could hardly be truly defined as such.

As it has been noted, moreover, “(t)he consensus from the critical smart cities literature is that little is known about the underlying principles of the smart city model beyond the advertising campaigns of IT companies and the self-promotion of cities that acquire “smart systems” [...] The term smart is frequently laid on top of these essential technologies as a marketing or rhetorical device to justify a range of political and economic interventions”. Like “big data”, the term “smart city” can be used to indicate a wide range of technologies, processes, and practices, and this definitional fuzziness can very well be used hide the value judgements underlying particular policy decisions. Surveillance in the smart city appears to be a prime example. The instrumentation of the built environment can potentially mean, 

194 “Since 1990, the world has seen an increased gathering of its population in urban areas. This trend is not new, but relentless and has been marked by a remarkable increase in the absolute numbers of urban dwellers—from a yearly average of 57 million between 1990-2000 to 77 million between 2010-2015. In 1990, 43 per cent (2.3 billion) of the world’s population lived in urban areas; by 2015, this had grown to 54 per cent (4 billion). The increase in urban population has not been evenly spread throughout the world. Different regions have seen their urban populations grow more quickly, or less quickly, although virtually no region of the world can report a decrease in urbanization” - UN-HABITAT, ‘Urbanization and Development: Emerging Futures - World Cities Report 2016’ (2016) 6.
195 Sadowski and Pasquale (n 140).
196 Even though ‘ghost towns’ might be a misnomer – see e.g. Wade Shepard, ‘An Update On China’s Largest Ghost City - What Ordos Kangbashi Is Like Today’ Forbes (19 April 2016).
197 See Nam and Pardo (n 113).
198 The interest of cities in their own branding and image has been highlighted e.g. by Harrison and Donnelly (n 124).
199 See Söderström, Paasche and Klauser (n 173).
200 See Sadowski and Pasquale (n 140).
201 Innovation being indeed one of the key concepts associated to the notion of smart city by a relevant number of sources, as reported in Nam and Pardo (n 113).
202 See Hollands (n 86).
203 Gaffney and Robertson (n 84) 4.
concretely, the deployment of wide-ranging, granular, and persistent surveillance devices; the positioning and use of those technologies has to be subject to ex ante scrutiny, both from a legal and an ethical perspective. The smart city narrative could be used, for example, to try to justify the deployment of an overreaching CCTV system that would better be analysed within the securitization paradigm204 rather than within the “smart, sustainable growth” one. Words have a meaning, but the “smart city” wording is highly contextual, dependent on the concrete setting of reference: the smart city topos can be used as a container in which to fit a narrative to push through, regardless of how “smart” the proposed initiatives actually are.

2.2.5 TELEOLOGY OF THE SMART CITY

The smart city’s purposes are extremely wide: sustainable growth and optimal resource management have a prominent spot in most smart cities’ agenda, as much as pollution control, data-driven policymaking, evidence-based decision making, participatory and inclusive governance, security and safety, economic growth, and many other policy goals. While being a high-level concept, a wide-ranging paradigm shift rather than simply a product or a technological solution, the smart city does however concretely require the instrumentation of the built environment – the deployment of specific technologies aiming to tackle (more or less) defined issues. the adoption of such technologies by a municipality requires the allotment of significant funds. Budgetary allocations need a reason: the deployment of ICT within the urban environment, necessary precondition for a city to be considered ‘smart’, is conditioned to the existence of an issue to be solved. Instrumenting a city is obviously a costly effort: the initial expenditure needs a justification.

Each city has its own set of issue: crime control, education, mobility, healthcare, surveillance... Unless a conurbation gets built ex novo, the ICT necessary to make a city ‘smart’ will be ideally deployed in order to solve the most pressing issue considering the budget available to the municipality. As it was repeatedly highlighted during several unstructured interviews carried on with a number of stakeholders operating in the urban science and technology domains, smart cities start with the deployment of ICT solutions in a geographically delimited area, which then get networked, and have applications built on top of them. The initial push is an issue the city faces, and that is thought to be solvable through ICT. The urban ICT’s hardware gets then chosen, installed, networked, and interconnected, and then specific applications get written for it. The resources cities have at their disposal are of course finite, and their expenditure conditioned by the priorities each municipality has: instrumenting the built environment is a gradual process involving a multiplicity of choices that are ultimately driven by what a city needs.

At the same time, the functionalities that a smart city can perform are conditioned by what kind of technologies are implemented within the conurbation’s territory. They depend from the sensors and actuators concretely deployed within the built environment. At the time of writing, for instance, several European cities are facing the necessity of developing the infrastructure necessary to host a growing number of refugees fleeing from their country of origin. The ICT they are building in order to deal with the aforementioned migratory fluxes – let us say, for instance, for the purposes of security, surveillance, and crime control – will likely

remain integrated with the built environment long after the refugee crisis has passed, as urban planners are bound to consider. Despite the flexibility of modern networked ICT architectures, which grant the possibility to run a multiplicity of applications on the same piece of hardware, the functionality originally intended partially binds its further use: the purpose for which the urban environment has initially been instrumented conditions the city’s development.

Each city is different, as the challenges it faces depend on an infinity of contextual elements. The instrumentation of the built environment does not escape this logic: each city becomes smarter in order to face and adapt to the challenges that decision-makers choose to prioritize. There is a teleological drive to the smart city’s development: the deployment of ICT solutions within the urban environment always has a purpose, it always answers to the specific necessity each city decides to tackle. Underlying the selection of those purposes, and the prioritization of the issues the ICT is deployed to tackle, one can demarcate the values driving that specific city’s urban governance.

Part of the literature examined often portrays an idealized version of the smart city, where all its objectives – the purposes whose achievement drives the instrumentation of the built environment – are synchronically pursued, no matter the fact that the values underlying an objective (e.g. security, pursued through surveillance) might conflict with the ones underlying another (inclusive participation to democratic processes, pursued through free assembly). Smart cities on the ground appear to be different, their resources constrained, their expenditures bound to the solution of specific issues driven by underlying values. There is no single model of smart city since each city has its own set of problems and is driven by its own set of (and balance between) values, and even when one decides to interpret the notion of smart city to its simplest extent – as the deployment of this or that ICT solution within the urban ecosystem – technology is inherently political, conditioned by its context of reference, which changes from city to city.

2.2.6 A JUNGLE OF STANDARDS

In order to demarcate the concept of smart city, and to generate its working definition, it seemed appropriate to look at whether standards provided guidance on how to frame or narrow down the notion of smart city. Ultimately, one of the main drivers of the research underlying this dissertation is to find out how to avoid the unreasonable application of data protection legislation, one of the legislative barriers that might hamper the development of smart cities: the literature review undertaken was thus initially keen to identify the categories of data without which smart cities would not work. It became rapidly clear, however, how a defined concept of smart city does not currently exist, and how the widespread, horizontal range of activities cities are involved with would have made the categorization of the data a smart city necessitates a gargantuan task, extraneous to the scope and disciplinary area of this thesis.

The instrumentation of cities as a trend has a horizontal impact on all the aspects of city life and urban management: from energy provisioning to environmental protection to security to

---

205 E.g. Caragliu, Del Bo and Nijkamp (n 17).
207 “The influences, technologies, ideas, and trends that converged to produce Smart Cities are as diverse and varied as cities themselves”: Gaffney and Robertson (n 84) 1.
fostering economic growth. The number of standards applicable to each facet of the smart city environment is accordingly impressive, and does not help much narrowing the notion down. In fact, the activity of the existing standardization bodies on issues connected with the smart city topic has been portrayed as overwhelming, due to the range of activities cities are routinely involved with, to the width and broadness of the notion of smart city, and, arguably, to the hype surrounding it, which got several standardization bodies on the bandwagon.

While the concept of smart city is still novel, furthermore, many of the tasks that cities perform are as old as civilization is. Therefore, in addition to the booming number of standards explicitly referencing the term “smart city”, one would also have to consider all the ones applicable to the urban environment in general, if relating to its instrumentation and datafication. Standards relating to infrastructures, supply chain, various categories of data and metadata, transport and logistics, key performance indicators (KPIs), networking, and environmental information systems, for instance, are all an integral part of that system of systems the smart city – as any city – is. As it has been considered, for example, standards and specifications for open data, are not explicitly labelled as smart city standards. Yet, many of them have been adopted by metropolises to address various issues related to government transparency and accountability, innovation and growth, and efficiency in the delivery of various services. Those open data standards and specifications seem to belong to the concept of smart city as much as the ones detailing organisational structures or environmental indicators.

Even within the standardization community, the understanding of the concept of smart city is fragmented and diverse, highly dependent on the standard setting body’s own interpretation of its scope. The European Commission’s “Rolling plan for ICT standardisation”, a document which lists the EU standardization policy’s top priorities, for instance, appears to interpret the notion of smart city quite restrictively, framing it in the context of environmentally sustainable growth. Several of the other areas identified by the Commission’s Rolling plan, however, can be argued as being crucial for modern urban environments: smart grids and smart metering, the IoT, cloud computing, are all heavily standardized technologies that are core to modern smart cities. On the other hand, a number of standard setting bodies are developing an


209 Amongst them, the European Committee for Standardization (CEN), the European Committee for Electrotechnical Standardization (CENELEC) and the European Telecommunications Standards Institute (ETSI), ITU’s Telecommunication Standardization Sector (ITU-T), the British Standards Institution (BSI), the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), and several others.

210 ISO, for example, has published or is currently working on the following standards that explicitly reference the term ‘smart’: ISO 37120 on sustainable development & resilience of communities – Indicators for city services & quality of life; ISO/TR 37150 on Smart community infrastructures – Review of existing activities relevant to metric; ISO 37101 Sustainable development and resilience of communities – Management systems – General principles and requirements; ISO 37102 on Sustainable development and resilience of communities – Vocabulary; ISO/TR 37121 Inventory and review of existing indicators on sustainable development and resilience in cities; ISO/TS 37151 on Smart community infrastructure metrics – General principles and requirements; ISO/TR 37152 on Smart community infrastructures – Common framework for development and operation.

211 Lauriault and others (n 180) 10.


213 See e.g. ITU-T Focus Group on Smart Sustainable Cities (n 89); BSI, ‘PD 8100:2015 - Smart Cities Overview – Guide’ (2015); ISO/IEC JTC1 (n 85).
extensive and inclusive view of the notion, whose scope appears to cover most standardisable aspects of modern urbanities.

Rather than narrowing down the notion of smart city, looking into standardization activities only highlights the notion’s complexity and plasticity. Even through several iterations of structured queries aiming at narrowing the field down, the number of standards and the width and breadth of the topics they cover – from environmental data collection to building automation to privacy and information security – still paint an overly broad and fuzzy picture. This however also highlights the fact that the smart city – while enabled by ICT – is more than the mere deployment of technology within the urban environment, but involves a holistic shift in urban governance processes and strategies. The smart city concept is generally portrayed as a paradigm change enabled by ICT, a transformation in urbanism, rather than a set of standardisable technologies or processes.

2.2.7 DEFINING SMART CITIES?

The previous sections discussed the fundamental elements and the main perspectives pertaining to the smart city environment, as resulting from the literature review undertaken, in an attempt of reaching an understanding of what a smart city is. They highlighted the features that can be argued as being central building blocks of the concept of smart city, underlining at the same time the many different constituencies that the notion has. The original goal was to find the boundaries of the notion of smart city, and the requirements for their development, as one of the research questions motivating this dissertation implied that the conflict between open data and data protection would have hampered the smart city’s development by reducing or eliminating the possibility to re-use personal data. As the previous sections highlighted, however, there is not a univocal definition, conception, or idea of what a smart city is and of the information it requires to function. This subsection assembles this thesis’ understanding of what smart cities are.

The starting point of this inquiry is ICT, as its employment and deployment within the urban environment is a *conditio sine qua non* to be able to frame a smart city as truly such. The due emphasis on the human and organisational aspects of the evolution of the smart city environment, which is often highlighted by the available literature and whose importance is explicitly recognized by this thesis as well, should not sway the reader’s attention from the focal element that allowed cities to become ‘smart’ in the first place – information and communication technologies, embedded within the urban environment, networked and interlinked with each other. It is ICT that enables the shift in organizational practices and the investment in human capital that characterizes a holistic and mature vision of the smart city concept. As this thesis’ concern is, one way or the other, the fair use of information within the

---

214 BSI and Imperial College London Consultants, ‘Mapping Smart City Standards Based on a Data Flow Model’ (2015).
215 See BSI and Imperial College London Consultants (n 136), where standards are categorized in technical, process-related, and strategic.
216 “The most fundamental requirement to enable the smart city vision is the large scale instrumentation of the city’s infrastructure, which includes structural, utility, transport, environmental, government and industrial manufacturing infrastructure with sensors, actuators, tag, and readers and other sensing devices [...] This underlying sensor fabric needs to be coupled with a large-scale deployment of high-speed network infrastructure that not only supports the exponential growth in the number of connected devices but also meets the specific needs of the various vertical industries”: Balakrishna (n 164) 224.
smart city ecosystem, the definition would need to include the fact that the smart city – the data-driven city – lies upon two specular modalities of information collection. On one hand, a top-down approach, where data is gathered by the instrumentation embedded within the urban fabric itself; on the other, a bottom-up one, where the city’s intelligence derives from the information flowing from the myriad of ubiquitous portable computing devices, smartphones, smart-watches, tablets, and so on, part of which Greenfield aptly dubbed “Everyware”.\textsuperscript{217}

The top-down approach and the bottom-up ones are not mutually exclusive: in fact, smart cities and data-driven urbanism appear to deeply rely on both,\textsuperscript{218} as convenience and availability dictate. The stage and depth at which the ICT is embedded within the urban environment matters as well. Cities are either built as ‘smart’ from the ground up, retrofitted with ICT with a varying degree of rapidity and permeation, or merely marketed as ‘smart’ due to the employment of sectorial technological solutions, lacking a holistic approach or a long-term vision. The literature review highlighted also a dualistic approach towards the concept of smart city: namely, a corporate narrative, according to which the smart city is already here, and it is by and large just a matter of investing a municipality’s budget in the right technologies, and a more academically focused approach, much more critical about the characteristics and the very existence of the notion of smart city, concerned with its rationalization and benchmarking, or with the development of high-level smart city architectures or technologies to be employed within a particular smart city domain.

It seems also necessary to point out a number of additional attributes proper of the smart city context, part of which, in the definitions reviewed, often goes by without specific mention, mostly due to the fact that they belong within the conceptualization of cities or to the one of ‘smart’ objects as such, rather than to the by-products of the combination of the two – the smart city. First, the smart city is a geographically, politically, and operationally limited area,\textsuperscript{219} which leads it to be – as every city is – neatly delimited, both from a geographical perspective and from a political/administrative one. Borders condition a smart city’s operations, and partly limit the scope of the technologies deployed within a specific geographical portion of the earth: the merging between the physical and digital world\textsuperscript{220} that has been brought along by the rise of the concept of Internet of Things (IoT) should not lead to overlook the fact that the smart city occupies a finite, delimited space on earth.

The ‘smart’ in the ‘smart city’ syntagm, moreover, means more than just ‘interconnected’, ‘networked’ or ‘instrumented with ICT’ – most cities already are, and a growing number of them is about to be. ‘Smart’, in our context, alludes to the ability to analyse the environment and to consequently self-regulate, with a varying degree of autonomy, in order to ultimately somehow shape human behaviour. It is a term that borders with the notion of Ambient Intelligence:\textsuperscript{221} a property, rather than a set of specific requirements,\textsuperscript{222} that defines the status of those things which, due to the convergence of ubiquitous computing, high-capacity

\begin{thebibliography}{100}
\bibitem{217} Greenfield (n 108).
\bibitem{218} See Walravens, Breuer and Ballon (n 14).
\bibitem{219} See e.g. Shelton, Zook and Wiig (n 169); Nam and Pardo (n 113).
\bibitem{220} See Greengard (n 80); Sarah Eskens and others, ‘Beyond Control: Exploratory Study on the Discourse in Silicon Valley about Consumer Privacy in the Internet of Things’ (2016); Lessig, \textit{Code and Other Laws of Cyberspace} (n 67); Lessig, \textit{Code Version 2.0} (n 67); Shah and Kesan (n 131).
\bibitem{221} See De Hert and others (n 107); Ahonen and others (n 107).
\bibitem{222} Ahonen and others (n 107) 4.
\end{thebibliography}
communication infrastructures, and data analytics and interfaces able to adapt to human behaviour, eventually reshaping future interactions.\textsuperscript{223} There is more to the notion of ‘smart’ than the mere interconnection and digitalization of the brick-and-mortar environment: it is largely about the technology’s adaptiveness, and about its ability to (semi) autonomously modify human behaviour.

Finally, smart cities and the implementation of the technologies underlying their functioning are purpose-driven,\textsuperscript{224} which is to be intended in two different ways. On one hand, from a high-level perspective, smart cities are meant to fulfil a broad set of goals;\textsuperscript{225} there indeed seem to be a stark teleological character to the notion of smart city. On the other hand, the high amount of expenditure and the time required to administrations and technology providers to implement the ICT solutions on which the smart city is based naturally leads to the fact that there needs to be a specific need to be solved in order to drive public sector decision-makers to tender for the deployment of the particular technologies whose interconnection and embedment within the built environment is necessary to make cities ‘smart’.

What is, then, the fabled ‘smart city’? The potpourri of traits and characteristics above hardly allows to define the concept precisely. The more one mulls about it, the more the term ‘smart city’ just seems either a mere narrative, or just a synonym for ‘city’ viewed in the context of the rampant datafication and digitalisation of society. The most fitting approach to defining smart cities is to see them as a socio-technical construct.\textsuperscript{226} Even though the technological aspect is a dominant element of the smart city narrative, the concept of smart city does not indicate any single technology or bundle of technologies, as sensors, cloud infrastructures, and big data analytics – roughly, the technological underpinnings of the smart city\textsuperscript{227} – are deployed in many other contexts. “Smart city technologies become smart city technologies only by association with the idea of the smart city and the narratives, logics, practices, and symbolism of which it is constituted”.\textsuperscript{228}

There does not seem to be a unitary vision of the smart city, nor a common set of requirements or benchmark differentiating ‘smart cities’ from just ‘cities’ – just a narrative tying technology and the problematics faced by modern conurbations. While defining what ‘smart city’ means might end up being a pointless exercise, and there is no list of requirements to be fulfilled for cities to become ‘smart’, the review undertaken allowed to enucleate a number of elements that seem to be foundational to any smart city. First and foremost, ‘smart cities’ are cities – rectius, smart city initiatives are initiatives of urban renewal or creation, involving particular metropolitan areas, and aimed at specific goals and targeting determinate problematics. Second, those initiatives involve the deployment of technologies or the analysis of information gathered through technologies already deployed: the smart city is, at its very core, the entwinement between technology and a conurbation. Third, the deployment of technology at urban scale has social and political consequences, as technology is not neutral, but rather expresses certain policies and regulates the behaviour of urban dwellers. The fourth and last

\textsuperscript{223} Ahonen and others (n 107) xxi.
\textsuperscript{224} Batty and others (n 10) 482.
\textsuperscript{225} See Caragliu, Del Bo and Nijkamp (n 17); Nesbitt (n 144).
\textsuperscript{227} Edwards (n 7).
\textsuperscript{228} Sadowski and Bendor (n 226) 2.
element is the underlying narrative: the concept of smart city is a tale, a story used to push a certain imaginary of how urban development is bound to unfold.

There is, then, the inevitable filter constituted by the observer. The research underlying this thesis is performed within a specific disciplinary framing (legal studies) and focus (EU law). I do not doubt that the syntagm ‘smart city’ might mean something specific in other fields, like urbanism, or surveillance studies. As a matter of fact, the literature review highlighted a plethora of different definition, requirements, and perspectives on the meaning of ‘smart city’, many of which clearly signal the advent of new, data-driven ways of building cities. From my point of view, however, the ‘smart city’ just seems the large-scale deployment of ICT within the built environment. That does not mean to be reductive: technology is not neutral, but rather inherently political, and has regulatory capacity. It means that the deployment of the range of technologies through which the built environment is instrumented, and with which the data necessary for urban development are gathered, is conditional to an underlying set of problematics, and to a corresponding set of choices.

It has been underlined how the rise of the concept of smart city is intrinsically linked to a crisis narrative, according to which smart cities are meant to solve issues like urban overpopulation, economic austerity and unemployment, and environmental concerns. It has also been highlighted how there is an undeniable tendency to tackle social problems through technological means. The development of smart cities – the deployment of the specific technologies with which the built environment is instrumented – is thus conditioned both by the issues those technologies are meant to solve, and by the choices made with regard to their specific implementation. I see the concept of ‘smart city’ mostly as an empty container, fit to suit a multiplicity of different corporate and governmental narratives. Ultimately, in the real world, the utopian and scholarly notion of smart city is substantiated by the choice and implementation of a specific set of technologies, as a response to definite issues. Such set of technologies, as technology does, is bound to interact with the social dimension of the environment in which it is deployed, generating consequences at a political and regulatory level. Those elements are what should arguably be kept in mind during any discussion involving the term ‘smart city’: the technologies used, the issues they mean to solve, the goal they aim at achieving, and the social, political, and regulatory consequences of their implementation. While the smart city label may be useful for descriptive purposes, e.g. to craft narratives of sustainability, urban renewal and data-driven decision-making, what ultimately matters is the technology chosen, the reasons underlying its employment, and the consequences of its deployment.

2.3 SMART CITIES, PRIVACY, AND DATA PROTECTION

This thesis claimed, so far, that there is not a unitary definition of smart city, and that its best characterisation is of a paradigm shift in urban governance and management, enabled by ICT developments, towards data-driven and evidence-based urban policymaking. It also

---

229 Sadowski and Bendor (n 226) 9 ss.
231 Winner (n 138); Woodrow Hartzog, Privacy’s Blueprint: The Battle to Control the Design of New Technologies (Harvard University Press 2018) 44.
underlined how the ICT that instruments the built environment actively shapes individual and collective behaviour according to its underlying set of values. The smart city, if ill-conceived or poorly scoped, is naturally threatening for individuals’ rights to privacy and data protection. The instrumentation of the built environment means the placement of an array of interconnected sensors, CCTV cameras, big data analytic platforms, cloud computing infrastructures, IoT devices – potentially very intrusive technologies. It also means the “datafication” of the built environment, and its visual representation through dashboards – activities that do not merely represent reality, but shape it on their own. The smart city is however bigger than the mere sum of its technological parts: it implies a holistic shift in urban governance and management, and pushes forth an anthropocentric view of the built environment’s development that is necessarily bound to take individuals’ privacy and data protection into account.

In the absence of a countervailing push, every technology that can be used for surveillance and control will be used for surveillance and control, regardless of its original function. The “datafication” of the built environment and the technologies enabling it can be engineered or used for such purposes, evading the checks and balances that legitimise those activities in a democratic society. The smart city’s design must therefore consider data protection as a right and as a value, since the regulatory capacity of the technologies, practices, and policies in which the concept unfolds shapes what we do and are, as individuals and as a collectivity.

The instrumentation of the built environment can threaten individuals’ rights to privacy and data protection to an unprecedented scale, scope, and granularity, and that thus those rights should be considered as a primary nonfunctional requirement in the design, development, and deployment of the technologies underlying the smart city. The objectives of the smart city are extremely multifaceted, and yet all somehow aiming at increasing its citizens’ general quality of life. For that, it is paramount for the technologies underlying the smart city environment to consider, from their very outset, the rights and interests their misuse might infringe. Amongst them, the ones to privacy and data protection are undeniably prominent.

2.3.1 DATAFICATION AND DATA PROTECTION

The right to data protection stems from the right to privacy. Traditionally, in the EU, the right to private life as protected under Article 8 of the European Convention of Human Rights (ECHR) was interpreted as covering the right to privacy, its scope extended to data protection by the case law of the European Court of Human Rights (ECtHR) and by the Council of Europe’s Convention 108. However, as attested by Articles 7 (“Respect for private and family life”) and 8 (“Respect for private life and correspondence”), data protection is a fundamental right.

232 See Hartzog (n 231).
233 Edwards (n 7); Finch and Tene (n 21); van Zoonen (n 139).
235 As opposed to functional requirements, which dictate the concrete functions a technology must have, nonfunctional requirements relate to the overarching values and ideals on which a technology is based.
237 Council of Europe, Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. ETS No.108, Strasbourg, 28/01/1981.
and 8 (“Protection of personal data”) of the Charter of Fundamental Rights of the EU (CFR),\textsuperscript{238} there is now a difference between privacy and personal data protection. Indeed, “privacy embodies a range of values that are only partially advanced by data protection”,\textsuperscript{239} The right to data protection has been framed both as integral to privacy – a subset of its norms – or, by more modern doctrine, as an entirely different right.\textsuperscript{240} While privacy relates to some qualitative requirements (e.g. legality, necessity, legitimacy, proportionality) deriving from the Court of Justice of the European Union (CJEU) and the ECtHR’s jurisprudence, data protection can be likened to a set of rules stemming from the best practices in (personal) data processing that arose along with the development and diffusion of computing.\textsuperscript{241} Through those rules, the regulator set up a legislative wireframe meant to ensure fairness in data processing operations – the same kind of activities on which the smart city ecosystem’s analytics are based.

Data protection stemmed from the advances in computing capabilities, which grew exponentially from the mainframe era on, and from the risks that became associated with them. Over the years, many scholars, professionals, tinkerers, and thinkers\textsuperscript{242} drew attention on the threats arising from the power and information asymmetries deriving from the capacity of some actors to store, process, and make sense of a quantity of information that was not conceivable only a few decades before. The debate around the interaction between computers and privacy gave rise to the notion of privacy as control over one’s personal information\textsuperscript{243}, a precursor to the right to personal data protection. Computers were novel, and potentially scary, considering the values upheld by the concept of privacy: it appeared sensible to constrain data processing with an amount of principles meant to ensure fairness within information processing activities. The right to data protection is a set of rules instrumental for the safeguard of all the rights and freedoms that can be dented by the power and information asymmetries running between controllers and data subjects.

In the modern smart city ecosystem, the power of technology and its regulatory capacity still warrant a system of checks and balances meant to curb information asymmetries. While computing capacity became distributed – from mainframes to personal computers to ubiquitous computing – real informational power remains arguably centralised by a network of private and public actors with unparalleled access to data and processing capacity. Technology corporations, social networks, internet providers, data brokers, and state administrations have – each in its own way – a processing capability and an information availability that allows them to exert a considerable power on individuals and on the social formations in which they assemble. The smart city is a prime example of a collection of technologies whose regulatory capacity warrants a cautionary approach,\textsuperscript{244} just as mainframe computing and databanks were.

Such an approach is now explicitly sanctioned by EU law: according to the General Data Protection Regulation (GDPR), individuals’ right to data protection must be considered ex ante.

---


\textsuperscript{240} Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 214.

\textsuperscript{241} See further in Ch. 4.

\textsuperscript{242} Alan F Westin, Privacy and Freedom (Athenaeum 1967); Arthur Raphael Miller, The Assault on Privacy: Computers, Data Banks, and Dossiers (University of Michigan Press 1971).

\textsuperscript{243} Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 27 ss.

from the design phase on, not as an afterthought but engineered within the technologies through which urban environments are being instrumented. As clarified by the GDPR’s recital 78, when developing, designing, and deploying any technology or service based on the processing of personal information, producers must consider individuals’ right to data protection, and make sure that the entities using the technologies or running the services are able to fulfil their obligations. Article 25 of the GDPR, “Data protection by design and by default”, mandates that – considering the state of the art, costs, nature, scope, context, and purposes of processing, and the related risks for individuals – the data controller must implement appropriate technical and organisational measures, designed to apply the GDPR’s principles. Article 25 of the GDPR, if read on its own, could be mistaken for a mere statement of principle. However, when considered in conjunction with e.g. the articles relating to administrative fines ex Art. 83, or to the security of the processing operations ex Article 32, its practical and concrete enforceability results clear.

The legislator explicitly mandates the inclusion of the values upheld by data protection within the design of information processing technologies and processes. It implicitly recognises the political character245 and regulatory capacity246 of the technologies on which the smart city narrative is based. The smart city embodies exactly the kind of area in which data protection by design is crucial: a stack of potentially highly intrusive technologies that inevitably inform and regulate citizens’ behaviour, and that has a profound impact on its social context. The instrumentation of the built environment is highly sensitive to the values on which its development is based, as smart as the design and governance of the ICT on which it runs.

2.3.2 VALUES AND DESIGN

It would be unfair to characterise the smart city’s “quest for a new utopia”247 just as a measly bundle of technologies.248 If the smart city revolution really is a paradigm shift where “investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel [...] a high quality of life, with a wise management of natural resources, through participatory governance”, 249 then the technologies underlying its functioning need to be informed to the same principles and values on which its theory rests. This thesis argues that data protection should be considered from the outset of the process of instrumenting the built environment – from the design phase on – as a nonfunctional requirement in the development of the smart city ecosystem. I ground my argument on the values that the smart city’s development is purportedly meant to uphold, and on the recent explicit introduction of the principles of data protection by design and by default within the EU legal framework.250

245 Winner (n 138).
246 See Hartzog (n 231).
247 Townsend (n 6).
249 Caragliu, Del Bo and Nijkamp (n 17) 50.
In the past three decades, there has arguably been a disciplinary shift within the fields of design and ethics, a convergence of interests that has led to see technology as bound to accommodate a range of human values. Design turned to ethics, and ethics to design, in what has been dubbed “The Design Turn in Applied Ethics”. Value-Sensitive Design (VSD) aims at embedding values in technology’s design. It assumes that values and norms can inform the things we build, and exhorts at taking into consideration in advance the ethics, regulatory capacity, and political value of technology. Data protection is instrumental to those values. Its violation, as a right and as a principle, has been linked to a variety of harms and abuses. Its balancing with opposing rights and values is often conflictual, each clash to be solved on a case-by-case basis. However, its essence and importance as both a human right and as a value is hardly questionable.

The smart city is a prime example of the ideological push towards embedding values in design. The city’s instrumentation aims at ensuring better living conditions to inhabitants, embedding certain values (e.g. safety, sustainability, efficiency) within the urban tissue. The ICT underlying the smart city can however enable a disproportionate level of tracking and surveillance, in the absence of a countervailing push. Data protection embodies that drive: a set of values, and a related right, that aim at ensuring the respect for individuals’ privacy, autonomy, and (informational) self-determination. Data protection by policy acts through the law as a regulatory instrument. Data protection by design embeds fair information processing within a technological artefact’s requirements, making its design sensitive to the values and rights data protection is meant to uphold, and exploiting the regulatory capacity of technology. Considering data protection as a nonfunctional requirement in the development of the ICT used to instrument the built environment is in line with the main narratives pertaining to the smart city’s purposes, and may eventually contribute to framing such a blurry notion.

2.3.2.1 Which values?
This section has so far argued that the technologies and processes underlying the smart city environment – a wide array of sensors and actuators, big data analytic platforms, and cloud computing infrastructures – are inherently threatening to individuals’ privacy and data protection. The design and implementation of those technologies and processes should thus consider their respect as a nonfunctional requirement, as explicitly sanctioned by law. The development of the smart city environment should be informed by the values underlying privacy and data protection. What are, precisely, those values?

One of the main tenets underlying this thesis, as the previous subsections might have let out, is a firm belief in the formal and substantial dichotomy between privacy and data protection. Chapter 4 contains an in-depth discussion of the separation between privacy and data

---

251 Manders-Huits and van den Hoven (n 141) 54.
254 See e.g. Solove (n 252); Ryan Calo, ‘The Boundaries of Privacy Harm’ (2011) 86 Indiana Law Journal 1131. On the industry whose entire business model is based on the exploitation of individuals’ personal data, see Wolfie Christl and Sarah Spięberrmann, Netwoks of Control (Facultas 2016); Wolfie Christl, Katharina Kopp and Patrick Urs Riechert, ‘Corporate Surveillance in Everyday Life’ (2017).
255 Edwards (n 7).
256 GDPR, Art. 25.
257 See Ch. 4.
protection from an eminently legal perspective, but some of its elements will be anticipated in this section, as the split between the two rights derives, first and foremost, from the fact that they gradually came to enshrine different (albeit connected) values.

3.2.1.1 Privacy values in design

Privacy is an extremely multifaceted concept, connected to a myriad of different values such as seclusion and intimacy, autonomy and self-determination, trust, and integrity, amongst many others. Several authors provide their own systematization of the ways in which the concept of privacy is substantiated. Solove, for instance, moves from the harmful activities through which privacy can be dented, and in his taxonomy identifies four basic groups of possibly detrimental deeds (information collection, information processing, information dissemination, and invasion) each of which is further divided into sub-categories. Koops et al. provide a typology, rather than a taxonomy, and categorize the concept of privacy into eight “primary ideal types of privacy” (bodily, spatial, communicational, proprietary, intellectual, decisional, associational, and behavioural privacy), with informational privacy as an overarching element, tangential to each type. Nissenbaum does away with framing privacy within a specific casing by arguing that the context in which information flows sets a series of relative norms, the violation of which leads to what is perceived as a breach of privacy. The range of conceptual frameworks from which to choose is truly manifold – which is the fittest?

This thesis frames the smart city as a socio-technical construct: mainly a narrative, but undeniably driven by an eminent technological underpinning – by the growing instrumentation and datafication of the built environment. Regardless of the policy agenda pushed through the chosen narrative, the design of the technologies underlying the smart city is arguably the most important factor to keep into consideration when analysing the instrumentation of the built environment and its impact on privacy. Hartzog’s Privacy Blueprint seems to be the most appropriate description of the privacy values on which the smart city’s development should be informed, as it is directly concerned with the interaction between the design of new technologies and individual privacy. That is the framework chosen to describe – mutatis

---


260 Solove (n 252). Solove’s article, spurred by the perceived indeterminacy of the notion of privacy, continues on the lines of the categorization developed in William Prosser, ‘Privacy’ (1960) 48 California Law Review 381.

261 “In devising a taxonomy, there are many different ways to go about carving up the landscape. I focus on the activities that invade privacy” - Solove (n 252) 485.

262 The groups of privacy harms identified by Solove are Information Collection (1), which is divided in the subgroups of Surveillance (1a) and Interrogation (1b); Information Processing (2), which is divided in the subgroups of Aggregation (2a), Identification (2b), Insecurity (2c), Secondary Use (2d), and Exclusion (2e); Information Dissemination (3), which is divided in the subgroups of Breach of Confidentiality (3a), Disclosure (3b), Exposure (3c) Increased Accessibility (3d), Blackmail (3e), Appropriation (3f), and Distortion (3g); and Invasion (4), which is divided in the subgroups of Intrusion (4a) and Decisional Interference (4b); Solove (n 252).


265 Sadowski and Bendor (n 226).

266 Hartzog (n 231).
mutandis\textsuperscript{267} — the privacy values\textsuperscript{268} on which the development of smart city technologies ought to be based.

Hartzog argues that the design of new technologies is crucial for privacy (and hence for data protection),\textsuperscript{269} and that the law should consider it more seriously than it currently does.\textsuperscript{270} The argument is not novel in itself; it resonates with prior literature on the politics of artefacts,\textsuperscript{271} the regulatory power of technology and architecture,\textsuperscript{272} and the role of design in creating ‘choice architectures’\textsuperscript{273} that set the context in which individuals make decisions, and thus drive human behaviour. It also fits within the well-established Privacy by Design movement,\textsuperscript{274} which has been advocating for hardcoding privacy in products and processes from the design phase on for several years.

What is most relevant in Hartzog’s work, for the purposes of this section, is the identification of the privacy values in design worth focusing on: trust, obscurity, and autonomy. The fitness of Privacy’s Blueprint framework, aside from the conceptual proximity between the design of new technologies and the technological underpinning of the smart city narrative, has also the merits of providing further specificity to the concept of Privacy by Design.\textsuperscript{275} Furthermore, it moves past the over-reliance on the traditional notions of secrecy and control, which has arguably been hindering the process of modernization of privacy,\textsuperscript{276} both as a right and as a value.

Hartzog identifies in the concept of trust, rather than in the one of control,\textsuperscript{277} one of the main values towards which the design of new technologies should strive.\textsuperscript{278} The notion of control

\textsuperscript{267} When discussing privacy as a right, however, I will continue to adhere to European law and jurisprudence, and hence on the dichotomy between the fundamental right to privacy and the one to data protection, which leads to a substantial difference in framing the values that ought to underlie smart city technologies’ development and implementation.

\textsuperscript{268} Privacy and data protection as rights are dealt with in Ch. 4.

\textsuperscript{269} Most non-European authors do not differentiate between privacy and data protection in their writings, which is a dichotomy that derives (mostly) from the European legal tradition, rather referring to ‘data privacy’ or ‘information privacy’. That is bound to give rise to lexical inconsistencies: in the European milieu, privacy and data protection are not the same fundamental right. Regardless, many of the arguments made regarding data/information privacy are applicable to data protection. On the topic, see Lee A Bygrave, \textit{Data Privacy Law: An International Perspective} (Oxford University Press 2014) 23–29.

\textsuperscript{270} Hartzog (n 231) 7.

\textsuperscript{271} Winner (n 138).

\textsuperscript{272} Lessig, \textit{Code and Other Laws of Cyberspace} (n 67); Lessig, \textit{Code Version 2.0} (n 67); Shah and Kesan (n 131).


\textsuperscript{274} Which begun in the ‘90s and is often credited to the seminal activities of the Information and Privacy Commissioner of Ontario, Canada — see Ann Cavoukian, ‘Privacy by Design in Law, Policy and Practice’ [2011] A white paper for regulators, decision-makers and policy-makers.

\textsuperscript{275} “(P)rivacy by design is not a specific technology or product but a systematic approach to designing any technology that embeds privacy into the underlying specifications or architecture”, which differentiates it from Privacy Enhancing Technologies (PETs), a related concept often used as a synonym, but that rather indicates “applications or tools with discrete goals that address a single dimension of privacy”: Ira S Rubinstein, ‘Regulating Privacy by Design’ (2011) 26 Berkeley Tech. LJ 1409, 1411–1412.

\textsuperscript{276} Hartzog (n 231) 95 ss.

\textsuperscript{277} Which has historically been connected to privacy, see e.g. Westin (n 242).

\textsuperscript{278} Hartzog (n 231) 97. See also Neil Richards and Woodrow Hartzog, ‘Taking Trust Seriously in Privacy Law’ (2015) 19 Stan. Tech. L. Rev. 431; Waldman (n 259). Trust and smart cities have also been linked in Esther Keymolen and Astrid Voorwinden, ‘Can We Negotiate? Trust and the Rule of Law in the Smart City Paradigm’ [2019] International Review of Law, Computers & Technology 1.
over one’s personal information has been heavily criticised as a core privacy value in that, by assuming that people can adequately make choices about the safeguard of their informational privacy, shifts a large extent to the burden to individuals.\(^\text{279}\) Conversely, the concept of trust\(^\text{280}\) switches the focus back upon the entities that process personal data. Hartzog’s blueprint substantiates and specifies the value of trust in the related concepts of discretion, honesty, protection, and loyalty. Discretion, which can be roughly defined as the quality of behaving in a way that avoids revealing contextually sensitive information, is understood as an implicit element of most trustworthy information relationships.\(^\text{281}\) Honesty, seen as a proactive, affirmative obligation, is also a foundational value in the development of trustworthy technologies, as “deceptive technologies erode trust”.\(^\text{282}\) Likewise, as some definitions of trust suggest,\(^\text{283}\) protection – security – is a paramount element of trustworthy information technologies.\(^\text{284}\) Finally, trustworthy technologies should be loyal; Hartzog suggests that information technologies should be designed to avoid ‘self-dealing’ at the expense of the data subject.\(^\text{285}\)

The second main privacy value Hartzog’s Privacy Blueprint advocates for is obscurity. Traditional conceptions of privacy have frequently been based on the idea of secrecy, and on the categorisation of information relationships through the lenses of a strict public/private dichotomy.\(^\text{286}\) Such a binary approach does not adapt well to the nuances of actual information relationships, which are better understood by reference to the degree of accessibility and understandability of a given information in a certain context. Obscurity should be seen as a core privacy value in the development of smart city technologies, which should be designed to allow for a certain level of unintelligibility or inaccessibility of the (privacy-sensitive) information they gather. Hartzog et al. have made a comprehensive case for conceptualising obscurity as a core privacy value,\(^\text{287}\) mostly as applied to an online context. They argue that the

\(^{279}\) Richards and Hartzog (n 278) 444. From a practical perspective, moreover, the notice and choice paradigm in which the notion of control is substantiated in privacy has been shown to be hardly workable in practice: Hartzog (n 231) 62 ss. See as an example Aleecia M McDonald and Lorrie Faith Cranor, ‘The Cost of Reading Privacy Policies’ (2008) 4 ISJLP 543. Understanding how each smart city technology actually works and what kind of information it gathers is likely unfeasible for the vast majority of individuals. Additionally, the inherent opacity and asymmetry between who runs those technologies and the subjects whose data is processed makes control by the latter is dependent on a concession from the former.

\(^{280}\) Trust generally connotates the situation of a subject that, with some degree of voluntariness, chooses to rely on something or someone, and has been defined in several different ways by different authors and disciplines. Waldman, for instance, frames trust by saying that it begins when knowledge ends: Ari Ezra Waldman, Privacy as Trust: Information Privacy for an Information Age (Cambridge University Press 2018). Similarly, Schneier states that “trust begins where security ends”: Bruce Schneier, Liars and Outliers: Enabling the Trust That Society Needs to Thrive (John Wiley & Sons 2012).

\(^{281}\) Hartzog (n 231) 100.


\(^{283}\) Schneier (n 280).

\(^{284}\) Hartzog (n 231) 104.

\(^{285}\) Hartzog (n 231) 105. See also Richards and Hartzog (n 278).


degree of obscurity afforded to individuals in a given context is a function of the degree to which their information can be found (search visibility) and accessed (access protection), of the degree of their identification or identifiability, and of the grade of clarity to which such information can be understood. The criteria of search visibility, accessibility, identifiability, and clarity seem to be equally appropriate parameters to analyse the degree of obscurity provided by smart city technologies, particularly as the merger between code and space—the rise of what Hildebrandt defines as ‘onlife’—is contributing in blurring the distinction between human behaviour online and offline.

Finally, the third privacy value smart city technologies should be designed for is individual autonomy. Seeing in control over one’s information the main underpinning of privacy might be a flawed assumption, as it practically implies informational self-management abilities that are not common amongst average individuals. The idea of protecting individual autonomy, nonetheless, is still the bedrock of many— if not most—privacy theories. Hartzog’s Privacy Blueprint has the merit of operationalising the broad and blurry notion of autonomy as a privacy value in the design of new technology by tying it to the values of trust and obscurity. The question of whether the design of a particular smart city technology safeguards individual autonomy can be answered by considering how that technology engenders trust, understood as a combination of discretion, honesty, protection, and loyalty, and the degree of obscurity it affords to individuals in the context of reference, seen as a function of their information’s visibility, accessibility, identifiability, and clarity.

2.3.2.1.2 Is there a separate value to data protection?

As it is specifically tailored to the context of the development and deployment of new technologies, Hartzog’s Blueprint seems to be a suitable framework to pinpoint the privacy values on which the smart city should be informed. Does the same set of values underlie data protection as well? While what has come to be known as data protection within the EU is often called ‘data privacy’ or ‘information privacy’ in other jurisdictions, and even if data protection does derive from the right to privacy, they are still formally distinct fundamental rights within the EU legal system. This thesis holds that privacy and data protection came to become substantially distinct fundamental rights, divergent in their essence and rationale—despite a number of overlaps that are bound to derive from their logical and historical connection. Accordingly, it also holds that the values underlying data protection should be framed as (partly) distinct from the values underlying privacy. Data protection’s essence and rationale as a right are outlined in Chapter 4, but what is the value data protection aims at safeguarding?

---

288 Kitchin and Dodge (n 10); Mireille Hildebrandt, **Smart Technologies and the End (s) of Law: Novel Entanglements of Law and Technology** (Edward Elgar Publishing 2015).
289 Hartzog (n 231) 116.
290 Bygrave, **Data Privacy Law: An International Perspective** (n 269) 23–29.
291 The Charter of Fundamental Rights of the European Union enshrines the right to privacy in Art. 7, and the one to personal data protection in Art. 8.
292 See Ch. 4.
293 The theory of data protection as a right on which this thesis is informed is contained in Chapter 4, which is devoted to data protection’s rationale and essence. The value underlying data protection is however anticipated in this section, as it signals this thesis’ point of departure from the value framework developed in Hartzog’s Privacy Blueprint, on which I rely to frame the privacy values that should underlie the smart city’s development.
Hartzog’s privacy values’ framework moves from a critique of the Fair Information Processing Principles, also known as Fair Information Practices/Principles (FIPPs or FIPs). The FIPPs are a set of high-level principles developed from the ’70s whose content ended up being a sort of common core of most data privacy and data protection regulatory instruments all over the world. In the US, the FIPPs are the central to the sectoral legislation covering what is generally defined as ‘data privacy’. In the EU, the FIPPs are the soul of the omnibus legislative framework substantiating the fundamental right to personal data protection, while the right to privacy – which is broader than mere data privacy – has a whole other regulatory structure. It would thus be reasonable to assume Hartzog’s Privacy Blueprint as a valid value framework to pinpoint the values underlying EU data protection as much as US data privacy – after all, they are both based on the FIPPs.

However, as the following chapters will underline, the gradual process of (quasi)constitutionalisation of the right to personal data protection within the EU multi-level fundamental rights framework, which culminated with the adoption of the 2000 Charter of Fundamental Rights led to the development (or, rather, to the recognition) of a value other than the ones underlying the right to privacy. That value, the following chapters argue, is the acknowledgement, at a societal level, of the need of a comprehensive system of checks and balances regulating personal data processing. The value of data protection is, in a way, in the very existence of a system of rules and norms applying to the processing of personal data, regardless of its connection with concepts like privacy, or the secrecy and confidentiality of the information processed.

Such a system of rules and norms substantiates data protection, but does not equate to data protection per se. At the moment, the value underlying data protection materialises itself through the principles listed in the FIPPs, demonstrating a certain degree of consistency across different jurisdictions. The FIPPs, or even data protection’s reliance of the FIPPs, may very well change in the future; the optimal way of regulating personal data processing is not set in stone, and very much depends on the social norms existing at a given point in time. The value underlying data protection – the societal will to regulate personal data processing as a specific activity by virtue of its fundamental role in the modern society – remains, regardless of how it is specifically implemented, be it through the FIPPs or any other set of principles that may arise

294 Hartzog (n 231).
295 There are different versions of the FIPPs, each of which varies slightly in wording and number of elements, on account of the FIPPs evolution over time. The original list, which originates from a 1973 report of the US Department of Health, Education and Welfare, Secretary's Advisory Committee on Automated Personal Data Systems (titled “Records, computers, and the Rights of Citizens”) is as follows: (1) There must be no personal data record-keeping systems whose very existence is secret. (2) There must be a way for a person to find out what information about the person is in a record and how it is used. (3) There must be a way for a person to prevent information about the person that was obtained for one purpose from being used or made available for other purposes without the person's consent. (4) There must be a way for a person to correct or amend a record of identifiable information about the person. (5) Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuses of the data. See generally Robert Gellman, ‘Fair Information Practices: A Basic History’ (2019).
296 See Ch. 4 for the FIPPs’ history and gradual development into the EU’s legislative framework.
297 On what makes EU data protection and omnibus regime, see Lynskey, The Foundations of EU Data Protection Law (n 19).
298 See F Fabbrini, Fundamental Rights in Europe: Challenges and Transformations in Comparative Perspective (Oxford University Press 2014).
in the future. That is where the EU-centric focus of this thesis forces a partial departure from Hartzog’s Privacy Blueprint\(^\text{299}\) and its critique of the FIPPs. The framing of the values of trust, obscurity, and autonomy sits well within the scope of data/information privacy in design. However, the materialisation of data protection as a fundamental right distinct from privacy is to be seen as symptomatic of the emergence of a novel value: a sort of proceduralism applied to the processing of personal data, rather than to institutional legitimacy, of which the FIPPs are the current expression and implementation. In addition to the values underlying data \textit{privacy} – trust, obscurity, and autonomy – the design of new technologies should also consider the value underlying data \textit{protection}: the societal posture that arose as a response to the affordances of personal data processing technologies.

The theorisation of the values underlying privacy and data protection is an inherently subjective exercise and there are, consequently, a few foreseeable objections to my argument. One might argue, for instance, that data protection does not have autonomous standing, and is rather a facet of the broader concept of privacy. However, despite the fact that privacy – the respect for private life – is intrinsically impossible to define exhaustively\(^\text{300}\) both as a right and as a concept, over-inflating its scope could have the detrimental effect of diluting its meaning and value. Conversely, one might also argue that the laws and regulation through which privacy is substantiated are steering more and more towards data protection-like mechanics,\(^\text{301}\) and that (FIPPs-based) data protection is merely how data privacy is regulated. Yet, narrowing down data privacy to data protection seems quite reductive – it condenses the much broader concept of private life to a set of checks and balances regulating the processing of personal data. Similarly, it can also be argued that the design values underlying privacy identified in Hartzog’s Privacy Blueprint can also be embedded in modernized versions of the FIPPs,\(^\text{302}\) further blurring the distinction between the values underlying privacy and data protection. The FIPPs substantiate data protection, but do not equate to its underlying value \textit{per se}, and have been changed and updated throughout data protection’s entire history and across jurisdictions. Furthermore, data protection is intrinsically instrumental in nature: it shelters and enables other rights and freedoms, including privacy, however defined.\(^\text{303}\)

What is, then, the particular value underlying the right to data protection, which fostered its emergence as a standalone fundamental right within the EU’s (quasi)constitutional framework, and that differentiates it from privacy? This thesis argues\(^\text{304}\) that the value of the broad array

---

\(^{299}\) Hartzog (n 231).

\(^{300}\) The ECtHR, for instance, left the right to the respect for private life, which encompasses the right to privacy, undefined in several occasions: see e.g. ECtHR, Pretty v. United Kingdom (2002) 35 ECHR, §61; ECtHR, Niemietz v. Germany (1992) 16 EHRR 97 §29; Peck v. United Kingdom (2003) 36 EHRR 41, §57. More recently, see ECtHR, Antović and Mirković v Montenegro (2017) ECHR 1068, §41; ECtHR, Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland [GC] (2017) App no. 931/13, §129.

\(^{301}\) As testified, for instance, by the recent mushrooming of privacy bills in the US at both federal and state level, which seem to be modelled, to a large extent, on the FIPPs and on EU data protection secondary legislation.


\(^{303}\) Hartzog’s Privacy Blueprint itself recognizes the validity of (some of) the FIPPs: Hartzog (n 231) 60.

\(^{304}\) Chapter 4 develops the underlying argument further.
of rights and principles in which data protection substantiates is in the choice of having a comprehensive regulatory system meant to channel the specific activity of personal data processing by virtue of its riskiness for individuals, while generally allowing it by virtue of its potential for society. In a way, the value underlying data protection resides in the fact that it steers society towards a future where the possibility to process freely personal data is curtailed, to a certain extent, and made fair and transparent, rather than prohibited tout court or allowed regardless of the circumstances.

A critique to the approach outlined above concerns practical indeterminacy; how can values be implemented in design, anyway? The following subsection clarifies how the values discussed above can be concretely translated in the design of new technologies, such as the ones constituting the technological underpinning of the ‘smart city’ narrative.

2.3.2.2 But how? Implementing Privacy and Data Protection in Design

The debate revolving about ethics and values in technology has been booming over the past years, and I must admit – this thesis sees it with a fair degree of scepticism. On the one hand, the lack of enforceability inherent to ethical norms does not pair well with the power asymmetry fostered by the technological systems of systems that regulate people’s lives, both online and offline – a distinction that has progressively been waning away. On the other hand, values are foggy and often relative to the context where they are expressed: Facebook has been branding itself as a privacy company—“we take your privacy seriously” – while maintaining a business model that is at the very antipodes of the values underlying privacy and data protection.

The corporate world seems to love flaunting adherence to ethical ‘obligations’ and corporate ‘values’, concessions whose practical significance depends on their whims and magnanimity, as opposed to the law, whose protective capacity does not rely on the agreement of the ones it binds. Ethics and values, without proper implementation, are much more akin to public relations tools than to regulatory devices. There are, however, practical ways to implement the values underlying privacy and data protection in the stack of processes and technologies of which ‘smart cities’ are made, so that the ‘soft’ regulatory capacity of ethics and values gets bolstered by the ‘hard’ power of technology. That is to say that even when the values upheld by the rights to privacy and data protection ring hollow on their own – and they often do – there are ways to implement them within the design of the technologies from which they are innately threatened.

A technical perspective on privacy and data protection by design and by default is outside of the scope of this dissertation, and of the author’s expertise; this section, nonetheless, attempts

---

305 Data protection might very well be an individual right, but the obligations it mandates to data controllers have a collective scope.

306 “In the digital world, we observe a massive power imbalance between data processing entities, which determine what and how data is processed, and the individuals whose data is at stake [...] when using a specific service, many individuals are often unaware of the data processing and its consequences. Moreover, the user’s subsequent control over the nature of the processing that happens to their personal data once it is given away is limited. Lastly, penalties for infringements of legal data protection obligations usually take effect only after the fact” - George Danezis and others, ‘Privacy and Data Protection by Design - from Policy to Engineering’ (European Union Agency for Network and Information Security 2014) 1.


308 At least in theory.
a cursory description of a number of high-level privacy by design strategies that can be used to implement the values underlying privacy and data protection in smart city technologies, of the tactics that can be used to pursue those strategies, and of some of the privacy design patterns and privacy-enhancing technologies (PETs) that have been deemed fit for such purpose. The scholarship considered identified, by reviewing the state of the art, eight distinctive strategies and a number of related tactics, divided in two different categories: data-oriented strategies on one hand, and process-oriented ones, on the other. Data-oriented strategies focus on (personal) data processing itself, and are eminently technical in nature, while process-oriented strategies are concerned with the processes surrounding the responsible handling of personal data, dealing with the organisational and procedural aspects relating to personal data processing.

As for the data-oriented approach, a first strategy is to minimise the personal data processed by the system, limiting it as much as possible in consideration of the processing’s purpose(s). Data minimization, as a privacy by design strategy, can be achieved by either collecting data about fewer people, or less data about particular individuals, depending on the system in question. Viable minimization tactics are selecting ex ante only the most relevant attributes or people, and – conversely – excluding them in advance: whitelisting and blacklisting people or attributes are intuitive minimization tools. Personal data processing can also be minimised by stripping some attributes from a broader set of data as soon as their processing is not necessary anymore, or by destroying them tout court when they stop being useful. A relevant privacy pattern could be for instance opting for attribute-based credentials (ABCs) rather than for an identity-based system when developing or deploying a smart city technology or application that requires users’ authentication – but not necessarily their identification.

Separating the processing of personal data, either logically or physically, is another data-oriented design strategy that, by enhancing transaction costs necessary to combine or correlate data, reduces the risk that information from one context gets shared or used in another context. The separation strategy can be achieved through two tactics: isolation – deciding to process personal data through separate (but still centralised) databases or applications – and distribution, where personal data are processed in a decentralised, distributed system. A pertinent privacy pattern is given by choosing local personal data

---

309 Tactics are an additional level of abstraction between design strategies on one hand and design patterns on the other; see Michael Colesky, Jaap-Henk Hoepman and Christiaan Hillen, ‘A Critical Analysis of Privacy Design Strategies’, 2016 IEEE Security and Privacy Workshops (SPW) (IEEE 2016). As it has been noted, “Privacy by design is a multifaceted concept: in legal documents on one hand, it is generally described in very broad terms as a general principle; by computer scientists and engineers on the other hand it is often equated with the use of specific privacy-enhancing technologies (PETs). However, privacy by design is neither a collection of mere general principles nor can it be reduced to the implementation of PETs. In fact, it is a process involving various technological and organisational components, which implement privacy and data protection principles” - Danezis and others (n 306) 3.


storage\textsuperscript{313} (e.g. on a mobile phone or on a USB key) rather than a centralised architecture: think e.g. of parking access – how much of an individual’s personal data does a parking garage need to store on its system, and what information can be stored on the user’s device instead?

Abstraction is a third strategy through which the values underlying privacy and data protection can be safeguarded by designing technologies as to limit, as much as possible in light of the processing purposes, the detail in which personal data is processed. As opposed to minimisation, abstraction does not limit the quantity of the data processed, but its granularity. Abstraction can be implemented by summarising certain attributes into less granular versions (e.g. by processing whether a data subject is a minor or not rather than the exact age), by grouping individual data into an aggregate group profile, or by perturbing the personal data processed by approximation or through the addition of random noise. A relevant pattern would be for instance location data fusing,\textsuperscript{314} through which the accuracy of location information is decreased in a way that aims at preserving its general utility. Let us say that a smart city project requires the monitoring of individuals’ location data to identify aggregated movement patterns to be used as a basis for data-driven spatial planning: to what extent can the information capture technologies deployed abstract the data gathered while still being able to fulfil their goals?

Even when personal data must be collected, and regardless of the level of detail necessary to achieve the processing’s purposes, data subjects’ information can be concealed. Hiding personal data, a last data-oriented strategy, can be achieved through several tactics: access restrictions, obfuscation (e.g. through encryption), disassociation to prevent linkability, and mixing it to hide its origin or its relationship with other data. The use of pseudonymous identities\textsuperscript{315} is a design pattern that well exemplifies how hiding personal information can be a valid design strategy – if a smart city service provider does not need to know a user’s real identity, it may very well be required to deal with a pseudonym, so that the identification of the data subject’s identity would need an additional step (and further transaction costs) to be performed. Likewise, when users’ inputs can lead to privacy breaches, a system that foresees the possibility to generate fake inputs that cannot be distinguished from real inputs so that the system operator cannot identify an unidentified data subject or infer the attributes of an identified data subject\textsuperscript{316} is a viable privacy pattern to implement the ‘hide’ strategy.

Aside from the data-oriented privacy by design strategies and tactics above, there are also a number of process-oriented options through which the values underlying privacy and data protection can be transposed into the technologies that make up the technological layer of the smart city construct. The provision of information by design helps curbing the information asymmetry inherent to the deployment of smart city technologies. Informing data subjects, timely and adequately, about the processing of their personal data is a first process-oriented strategy, and can be implemented through the tactics of supplying information about the processing to data subjects, explaining its logic, and notifying individual data subjects about processing events that specifically concern them. Examples of design patterns that implement

\begin{itemize}
\item \textsuperscript{313} See https://privacypatterns.org/patterns/Personal-data-store.
\item \textsuperscript{314} See https://privacypatterns.org/patterns/Location-granularity.
\item \textsuperscript{315} See https://privacypatterns.org/patterns/Pseudonymous-identity.
\item \textsuperscript{316} See https://privacypatterns.org/patterns/Use-of-dummies.
\end{itemize}
the ‘inform’ design strategy are privacy dashboards, through which data subjects can control how a data controller processes their information, and privacy mirrors, which provide data subjects with a reflection of how a certain system processes which personal data about them.

Information asymmetries come with, or lead to, power asymmetries as well. Designing personal data processing technologies and processes to allow for some degree of individual control over the processing is another viable design strategy. At a lower level of abstraction, i.e. having regard to the relevant tactics, smart city technologies can be designed to foster individual autonomy by requiring, when appropriate, data subjects’ informed consent to the processing of personal data, allowing them to choose an alternative to the processing if possible (e.g. offering an alternative to smart metering). The technologies underlying the smart city environment can also be designed to provide individuals with the means to review, update and retract their personal data, for instance through a privacy dashboard. A design pattern implementing the ‘control’ strategy could be supporting selective disclosure controls, the possibility to decide ex ante with whom certain personal data are to be shared, and for what purposes – a data subject might accept to allow its energy provider to use its smart meter data for research purposes, while another might not.

The processes underlying personal data processing in the smart city environment can only be privacy and data protection-friendly if there is an adequate degree of (self)enforcement by the data controller. Enforcement, as a data protection by design strategy, can be implement through the tactics of creating and enforcing a specified privacy policy, maintaining it through the necessary technical and organisational measures, and upholding it as circumstances change. Sticky policies – machine-readable policies attached to (personal) data to define the usage permitted – are an example of a design pattern implementing the ‘enforce strategy.

As good privacy and data protection practices require accountability from data controllers, a last process-oriented privacy by design strategy is to demonstrate that personal data is processed as fairly as specified in the relevant policies through the associated tactics of recording logs relating to the processing activities performed, to audit those logs and the underlying processes, and to report the results of the logs’ audits to the relevant stakeholders. Automatic event logging exemplifies the kind of pattern that can substantiate the ‘demonstrate’ privacy by design strategy.

Implementing values in design does not necessarily have to be just ethics-washing: there are ways to embed privacy and data protection in the development and deployment of the technologies underlying the smart city environment. Smart city technologies – smart meters, local e-government services, automatic license plate readers, and the likes – are all data capture technologies, and hence inherently threatening to individuals’ privacy and data.

---

317 For a review of relevant patterns, see e.g. Michael Colesky and Julio C Caiza, ‘A System of Privacy Patterns for Informing Users: Creating a Pattern System’, Proceedings of the 23rd European Conference on Pattern Languages of Programs (ACM 2018).
321 For a quick and convenient list of smart city technologies, see Future of Privacy Forum, ‘Shedding Light on Smart City Privacy’ <https://fpf.org/2017/03/30/smart-cities/>.
protection, but it is nonetheless possible to develop and deploy them in a way that makes the intrusion proportionate to the goals to be achieved. Depending on the political will and on the resources available to a city’s administration, however organised, it is possible to balance the values underlying privacy – trust, obscurity, and autonomy – and the set of checks and balances that substantiates data protection with the inevitable instrumentation and datafication of the modern built environment.

2.4 SMART CITIES AND OPEN DATA

The connection between open data – particularly as directed at fostering economic growth and public sector efficiency – and the smart city environment should result apparent from the sections above. The re-use of PSI and the instrumentation of the built environment have largely overlapping goals: generating value through the exploitation of information, be it data already held by the public sector or data gathered through the distributed sensors’ network characterizing smart city infrastructures. The connection between open data and the success of future smart city initiatives is also intuitively bound to increase as the availability, granularity, kinds, and overall quality of the existing data keeps growing; it can be roughly portrayed as a sort of feedback loop, in which the information gathered by the public bodies or the public-private partnerships running the smart city infrastructure is released as open data and used as a basis for data-driven urbanism.

Open data can be therefore seen as a cornerstone for the development of smart city initiatives, enabling, amongst other things, the creation of new business models, the rationalization of public sector activities to gain in efficiency and savings, and transparency and inclusion in governance and decision-making processes. Open data can act as a catalyst, a connector bringing together different levels of government and the other relevant stakeholders (industry, academia, civil society, individuals) within the context of the smart city ecosystem. Open data can be used to generate economic growth, and to make cities better.

On the other hand, open data is a by-product of smart city initiatives as well: the rampant instrumentation of the built environment – and its consequences on the human and organizational components of the urban ecosystem – naturally leads to an increase in the availability of information and related services, part of which is bound to be released in an open format. The array of sensors with which the environment is being measured and ‘datafied’ and that characterizes many understandings of the ‘smart city’ buzzword will arguably be as important to future open data availability as the handiness of granular and high-quality information is to modern-day urbanism.

Be it as it may, one of the research questions underlying this thesis inquires whether the purported conflict between open data and personal data protection hampers the development of smart cities. In consideration of the review undertaken in the previous sections of this chapter, I believe that such question cannot be answered unless tied to a specific city, or rather to a particular initiative concerning a certain city, as ‘smart city’ is a mere umbrella term, devoid

322 Finch and Tene (n 21); van Zoonen (n 139).
324 Kitchin, ‘The Real-Time City? Big Data and Smart Urbanism’ (n 91); Kitchin, ‘Data-Driven, Networked Urbanism’ (n 3); Kitchin and Dodge (n 10).
325 Walravens, Breuer and Ballon (n 14).
of specific requirements. Regardless of the interaction between open data and data protection, which will be explored in the rest of this thesis, each smart city initiative is different, and has its own goals, requirements, and limits.

It may very well be that open data is foundational to the vision underlying a particular smart city initiative, perhaps one aiming at bolstering the city’s digital economy by fostering public-sector information re-use by private actors, while being absolutely negligible for another one, maybe one aiming at reducing crime through the deployment of a sophisticated CCTV network. The Sidewalk Toronto consortium, for instance, is considering to create an open data hub, based on standard formats and interfaces, to “allow for faster innovation”, creating an open data innovation ecosystem while at the same time avoiding vendor lock-in. Sidewalk Labs plans to do so by supporting existing standards where they exist, and develop new formats and Application Programming Interfaces (APIs) (in collaboration with the relevant stakeholders) where they do not. The aim seems for Quayside to provide an environment similar to Internet, where developers get to use well documented APIs, so that “the rate of technology innovation will increase substantially”.

The municipality of Barcelona is also pursuing ‘smartness’ through (inter alia) open data infrastructures, albeit following a more communitarian ethos than Sidewalk Toronto’s neoliberal approach.

What ‘smart city’ means in practice depends on the technology deployed, the issues to solve and the goals to achieve, and the social consequences of the city’s instrumentation. Every other element surrounding the notion seems to pertain to a “narrative about urban crises and technological salvation”, one meant to fit different ideas and initiatives into a coherent view of smart urbanism, and disseminate a specific vision of smartness while crowding out alternative visions of ‘smart city’. Tying the notion of smart city to the concept of open data is fairly frictionless. Constructing a smart city narrative that glorifies open data as core to the instrumentation of the built environment is a mere matter of framing. Open data can be portrayed as an important by-product of a particular smart city initiative, thus framing it as one of the benefits deriving from it. Conversely, open data can also be framed as a foundational element of data-driven urbanism which often seems to be synonymous with the notion of ‘smart city’. While information tout court is certainly essential to any smart city vision, however, the same cannot be said of open data – the relevant stakeholders can very well choose to keep the information they gather somewhat closed (unless otherwise prescribed by law) and still operate in a ‘smart city’, or just brand it as such.

In the end, what really matters is the design and implementation of the technologies and processes that instrument the built environment and make a specific area ‘smart’. The presence or absence of an open data infrastructure underlying a certain smart city vision, in

326 See e.g. Open North (n 33); Lauriault and others (n 180); Tracey Lauriault, Rachel Bloom and Jean-Noé Landry, ‘Open Smart Cities in Canada: Assessment Report’ (2018).
327 See e.g. Nora Ni Loideain, ‘Cape Town as a Smart and Safe City: Implications for Governance and Data Privacy’ (2017) 7 International Data Privacy Law 314; Gaffney and Robertson (n 84).
328 https://sidewalktoronto.ca.
330 Sidewalk Labs (n 329).
331 See Ajuntament de Barcelona, ‘Pla Digital de l’ Ajuntament de Barcelona Mesura de Gestió Ètica i Responsable de Dades: Barcelona Data Commons’.
332 Sadowski and Bendor (n 226).
333 E.g. by PSI legislation; see the next chapter.
the absence of countervailing legal requirements, depends on the existence of a political will driving its existence. The question of whether the purported\(^{334}\) conflict between open data and data protection hampers the development of the concept of smart city, which underlies this thesis, is essentially flawed to the extent that it assumes that open data is inherently a requirement for smart cities. In other words, it is impossible to say to what extent, if any, the obstacle to open data sharing constituted by data protection law hampers the development of the concept of smart city, as there are no specific, univocal requirements to be fulfilled for cities to be defined as ‘smart’.

To be clear, information processing is undoubtedly foundational to the development of the smart city environment – whatever that may turn out to be. Open data, as the following chapter aims at highlighting, is not however synonymous with data sharing, and its inclusion within a particular smart city narrative is not naturally inherent to the notion of smart city per se, but rather dependant on the policy agenda underlying the crafting of the definition in question. For example, Open North, a Canadian non-profit organisation, developed a rather long and convoluted smart city definition,\(^{335}\) very aspirational in nature and centred on openness, which sees open data as a core element of the smart city concept. Regardless of how embraceable the elements of Open North’s Open Smart City definition are, they do not really define smart city ‘requirements’, but rather push forth a policy agenda, as other definitions do. Is data protection legislation a barrier to open data sharing, and would such conflict hamper the development of the concept of ‘smart city’, as one of this thesis’ research questions postulates?\(^{336}\) Such question pushes forth a specific agenda by framing open data as a smart city requirement, thus tacitly moving from a pre-conceived, implicit definition of what a ‘smart city’ is, which does not necessarily reflect other definitions\(^{337}\) or real-life implementations\(^{338}\) of the concept.

The notion of smart city is a socio-technical construct,\(^{339}\) a narrative underpinned by a wide and varied stack of technologies, which does not have ‘requirements’ that are made unattainable by the conflict between open data and data protection; the mere reference to the smart city, in other words, does not provide a reason to alter the balance between open data and data protection. Regardless of whether a city is branded as smart or not, which is devoid of legal consequences, European municipalities and the public bodies operating within them are still subject to EU and national law mandating (open) data sharing, such as the PSI Directive\(^{340}\) and its implementation in Member States’ legislation. The following chapters will

---

\(^{334}\) As the following chapters will highlight, 1. There are various degrees of ‘openness’ within the notion of open data as generally understood 2. Data protection is a permissive regime by design 3.

\(^{335}\) Open North (n 33).

\(^{336}\) Open North’s smart city definition itself frames the respect for individuals’ privacy and data protection as a fundamental requirement in the development of the ‘open smart city’: Open North (n 33) 25 ss.

\(^{337}\) See (amongst many other collections of smart city definitions) Cocchia (n 9); Farnaz Mosannenzadeh and Daniele Vettorato, ‘Defining Smart City. A Conceptual Framework Based on Keyword Analysis’ [2014] TeMA Journal of Land Use, Mobility and Environment; Albino, Berardi and Dangelico (n 75); ITU- T Focus Group on Smart Sustainable Cities (n 89).

\(^{338}\) E.g. Rio: see Gaffney and Robertson (n 84).

\(^{339}\) Sadowski and Bendor (n 226).

hence focus on the relationship between the law and policies promoting open data, on one hand, and the right to data protection, on the other, irrespective of open data’s framing as a requirement for the smart city’s development. The smart city topos will still be used as a source of examples of the rampant instrumentation of society, and of its effects on the balancing between open data and data protection.

2.5 CONCLUSION

Cities are complex and multifaceted entities, convoluted systems of systems that constantly adapt to the challenges and needs of the communities they serve. It has been considered for several decades how cities cannot be reduced to a set of metrics and indicators, and that they develop organically in ways that are not always foreseen by traditional, orthodox urban planning theory.341 There is, however, an undeniable tendency towards the instrumentation and datafication of cities, and a drive towards data-driven decisions and policy-making, which has been identified and defined by reference to the wording ‘smart city’.

A cursory analysis of the available academic, corporate, and governmental literature on the topic has highlighted a plethora of perspective and approaches to what a ‘smart city’ is supposed to be. All in all, ‘smart cities’ seem to be just cities whose problematics are tackled through the deployment of particular ICT solutions – smart meters, a wide array of sensors, automatic license plate readers, and so on – which are often integrated through some sort of control room or city dashboard. In addition to that, the fabled ‘smart city’ is also a narrative through which a plethora of varied stakeholders attempt at influencing the public discourse about how cities are shaped and governed.

Nonetheless, the technology deployed to tackle a city’s challenges is not neutral, but rather inherently political, and has regulatory capacity too. The values on which the design and deployment of the technologies instrumenting the city are based shape how modern urbanities develop and how individuals live them. At the root of the instrumentation and datafication of the urban environment lies a certain set of problematics – environmental concerns and unemployment, traffic congestions and spiking crime rates – whose solution concretely depends on the choices made by the relevant stakeholders and on the values driving them. What each ‘smart city’ is, concretely, depends on those choices and on those values: there is not a single model or archetype of what makes a city ‘smart’, nor a standard set of requirements to fulfil.

Individuals’ privacy and data protection are naturally threatened by the technologies on which the smart city is based. Those technologies are intrinsically data capture technologies, and revolve on the processing of information. Some of that information is bound to be personal data – information relating to an identified or identifiable natural person; an even larger amount of data will undoubtedly relate to groups, provoking issues about group privacy342 that

---


342 On the topic of group privacy, see Linnet Taylor, Luciano Floridi and Bart van der Sloot, Group Privacy: New Challenges of Data Technologies, vol 126 (Springer 2016).
are not directly addressed by current data protection legislation.\textsuperscript{343} There are, however, ways to embed the values underlying privacy and data protection in the design and in the deployment of the technologies instrumenting the built environment.

Indeed, data processing and information sharing seem to be foundational to the concept of smart city, regardless of the values and narratives underpinning it – what else would the ‘smart’ in ‘smart city’ signify, otherwise? That does not necessarily mean, however, that open data – a particular data sharing policy – is a requirement for the development of smart cities. One of the postulates from which this thesis moved from is that data protection legislation, by reducing the possibility to share personal data as open data, could have the detrimental effect of hampering the development of the smart city as a concept and of smart cities in the real world. That is, however, a misleading assumption, as there is no univocal, authoritative, or even agreed-upon definition of what a smart city should be, nor of its requirements or of the kinds of data it needs. The idea of ‘smart city’ is rather a white canvas, an aspirational concept fit to accommodate a plethora of different narratives and agendas, where open data may very well be of the utmost prominence, but also completely negligible. It all depends on the policy agenda underlying the instrumentation of the urban environment; what really matter are the problematics and challenges a city faces, the technologies it chooses in order to tackle them, and the values underlying the design and deployment of those technologies.

The tension between open data and data protection – the main focus of this thesis – is surely exacerbated by the rampant instrumentation and datafication of the urban environment. Yet, the introduction of the smart city topos in this (policy) debate is immaterial to the balance between open data and individuals’ rights to privacy and data protection. The smart city narrative does not alter the relationship and balance between open data laws and policies, on one hand, and the rights to privacy and data protection, on the other hand. Data protection is meant to be technology-neutral:\textsuperscript{344} it would be unfeasible to update data protection law as soon as a new technological development involving personal data processing comes to light. Big data analytics, cloud computing, machine learning and artificial intelligence, and distributed ledger technologies are all innovations that somehow challenge the application of data protection law, which has historically been based on principles developed when the dominant (rectius: only) computing architectures were centralised,\textsuperscript{345} but that have always been dealt with through the existing legal framework,\textsuperscript{346} one way or another. That holds even more true in consideration of the fact that the smart city is not (yet) a specific technology or

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{343} See Alessandro Mantelero, ‘Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection’ (2016) 32 Computer Law & Security Review 238.
\item \textsuperscript{345} See Ceruzzi, \textit{A History of Modern Computing} (n 99); Ceruzzi, \textit{Computing: A Concise History} (n 99).
\item \textsuperscript{346} For instance, see how the cloud computing paradigm, which arguably revolutionised modern computing (and, hence, modern society), was dealt with within the boundaries of the existing regulatory regimes in Christopher Millard, \textit{Cloud Computing Law} (Oxford University Press 2013). The introduction of new technologies, however revolutionary, seems to seldom warrant the overhaul of the applicable legal frameworks, at least on their own.
\end{itemize}
\end{footnotesize}
technological paradigm, like cloud computing or blockchain ledgers, but rather a narrative, a socio-technical construct based on the deployment of a (very wide) set of technologies meant to address a varied array of problems in a broad assortment of contexts. That is not to say, again, that there is no conflict between open data laws and policies and EU data protection legislation, but that the smart city narrative does not shift the balance between them. The relationship between open data and data protection remains the same regardless of whether one considers it in light of the fabled ‘smart city’ environment or not.

The narrative of technological salvation that underpins the concept of ‘smart city’ can be seamlessly transposed to a multiplicity of different contexts – smart healthcare, smart transportation, smart government, and so forth – without any substantial alteration of sort. The inclusion of open data within the ‘requirements’ underlying the smart city’s development does not derive from it being a conditio sine qua non for cities to become ‘smart’ (whatever ‘smart’ might mean), but rather from a specific policy agenda – an amenable one, one of innovation and growth through data interoperability and re-use, but an agenda nonetheless. Arguing that the purported conflict between open data and data protection would hamper the development of the ‘smart city’ utopia is thus artificial, and – as the following chapters will show – tantamount to arguing for the deregulation of the activity of personal data processing.

The following chapter deals with open data by reconstructing the origins of the concept, and its peculiarities vis-à-vis ‘traditional’ data sharing: the term ‘open data’ is often used to indicate a wide array of somewhat permissive data sharing policies, but not all of those policies actually integrate the requirements for information to be considered as actual open data. It will then delve into (some of) the legislative instruments implementing open data policies within the EU, and in particular into the PSI Directive, which has recently been updated and recast. The chapter will argue that, while data protection is an autonomous fundamental right within the EU (quasi)constitutional framework, open data is merely a data sharing policy, which can uphold (directly or indirectly) fundamental rights, but is not a right on its own. The following chapters will then deal with the right to data protection – what it is, how it emerged, and what are its essence and substance – and on interaction with open data at a primary and then at a secondary level.

347 See Mell and Grance (n 109).
348 See Dylan Yaga and others, ‘NISTIR 8202 Blockchain Technology Overview’ (National Institute of Standards and Technology 2018).
349 Sadowski and Bendor (n 226).
350 See Future of Privacy Forum (n 321).
351 See Morozov (n 230).
3. Open Data

As code permeates the built environment, intertwining with the infrastructures and the spatialities of the urban ecosystem of reference, information becomes an essential element in a city’s development. While the datafication of cities is undeniably an activity that has far reaching roots – cartography being clear evidence of that – the instrumentation of urban spaces and the parallel ubiquity of mobile computing devices reshaped the relationship that cities have with data, inflating its importance and its prominence within urban decision-making processes.

Part of the data employed by and used within the smart city environment is of course bound to be closed, being either confidential, sensitive, proprietary, or a combination of the above, and also due to the rise of public-private partnerships and to the rampant privatization of functions traditionally assigned to a city’s governance, which naturally clashes with the openness required from public administrations. An increasing amount of information, both with regard to its sheer size and to its proportion with closed data, is however being released as open data within the urban context.

Regarding the smart city environment, this thesis postulates that the role that open data is bound to play can be framed as twofold: on one hand, open data can be seen as one of the smart city’s cornerstones, a precondition for truly ‘smart’ environments to function as such, and therefore to increase transparency, efficiency, and growth. On the other, the sensing and processing capabilities that characterize the smart city environment allow for an ever-increasing potential for data collection, sharing, and reuse, thus potentially boost the amount, kinds, quality, and granularity of open datasets.

This chapter therefore begins by elucidating the concept of open data as presently conceived, and by giving an historical and genealogical account of its recent evolution. It then delves into the different rationalities for open data, pointing out the distinctive objectives that open data instances aim to achieve, and highlighting how those objectives are pursued through EU legislation.

3.1 What is Open Data?

The term ‘open data’ generally refers to the concept that (some kinds of) information should be completely free and open to all kinds of use, re-use, and distribution, by anyone and for any purpose, and that such freedom should be ensured through both legal and technical means. Such idea is substantiated, concretely, in a set of requirements that relate to the information and/or to the license covering it. That is the main connotation of the wording ‘open data’ used in this dissertation.

At a higher level, however, ‘open data’ may also refer to institutional policies, adopted either spontaneously or as mandated by applicable legislation, meant to foster data accessibility, usability, re-usability, and distribution. Moreover, ‘open data’ sometimes also indicates (one

---

352 See Kitchin and Dodge (n 10).
353 E.g. the PSI Directive.
of) the movement(s) that have been pushing the general idea that ‘information wants to be free’\textsuperscript{354} with particular respect to opening up (governmental) data. Open data and its requirements have been defined in slightly different ways by several actors and stakeholders. Yet, all the definitions encountered during this research, despite their differences in wording and detail, present roughly the same set of conditions that must be respected to qualify a dataset as open data.\textsuperscript{355}

### 3.1.1 Open Data Requirements

According to the Open Definition,\textsuperscript{356} perhaps the most widely cited and accepted definitions available,\textsuperscript{357} open data can be defined as “data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and share-alike”. “Open” therefore refers to information that “anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)”.\textsuperscript{358} To be ‘open data’, a dataset must be either in the public domain or distributed through an open license. Furthermore, it must be provided as a whole, at reproduction cost at most, preferably through the internet and without charge. Openness is both technical and legal: information can be called open data only if it is provided in a readily machine-readable form, and the format in which the data is to be provided must be open as well, meaning that it does not place any restriction upon its use, and that files in that format can be processed with open-source software tools.

The Open Definition speaks broadly of open ‘works’, rather than of open data in particular. Another set of requirements that is often referred to are the Open Government Data (OGD)\textsuperscript{359} principles, which are the following:

1. **Data must be complete:**
   
   Open data must be the default setting for governmental data releases, provided that said data is of a public nature, meaning that it must not be subject to valid privacy, security or other legitimate and legally sanctioned limitations. Moreover, while some resources are by now digital by default, some other artefacts are not: in the latter case, they should be made digitally available to the maximum extent possible.

2. **Data must be primary:**


\textsuperscript{355}See below. This thesis, for convenience, moves mostly from the Open Knowledge Foundation’s Open Definition and from the requirements it sets, as it they are often understood to be the reference standard to define open data: see Open Knowledge International (n 32). The thesis however also considered a large number of other definitions, and distilled from them its own set of requirements, which are outlined below.

\textsuperscript{356}Open Knowledge International (n 32).

\textsuperscript{357}Notably, the 2019 recast of the PSI Directive, in its Recital 16, clearly leans on the Open Definition, albeit not adopting it explicitly: “Open data as a concept is generally understood to denote data in an open format that can be freely used, re-used and shared by anyone for any purpose.”


\textsuperscript{359}As pointed out by the principles’ authors, “(t)here are many definitions of “open” and this is but one. The 2007 working group’s definition sits at the unique intersection of open government and open data and has United States sensibilities” – Joshua Tauberer, ‘The Annotated 8 Principles of Open Government Data’ (2007) <https://opengovdata.org/> accessed 14 June 2016.
I.e. collected at its very source, and as granular as possible; entities obtaining the original dataset, processing it, and publishing it in a modified (e.g. aggregated) form, should have the obligation to publish the original data set in its default format, thus contributing in preserving it for posterity.

3. Data must be timely:
   Each piece of information has its own lifecycle, and the accuracy – and therefore the utility – of a dataset partly depends on the time lapsed from the dataset’s creation. Timeliness, therefore, means that government data should be released as early as possible, in order to preserve the data’s value.

4. Data must be accessible:
   Data must be accessible to the widest possible number of users for the widest possible array of purposes. In particular, accessibility is deemed to be lacking if the data is not accessible through automated means, due to technological, policy, or other kinds of restrictions.

5. Data must be machine-processable:
   As following from the accessibility principle above, data must be machine-readable and processable, and thus in a widely used, normalized and sufficiently documented format.

6. Data must be non-discriminatory:
   I.e. available to anyone, for any purpose, without access controls, ‘walled gardens’, use restrictions, or other gatekeeping activities that might lead to differences in accessing information.

7. Data must be non-proprietary:
   No entity should have exclusive control over the data format employed. As several proprietary formats are widely used, and conversely some open formats have a narrow users’ base, the decision of releasing a dataset in both widely used proprietary formats and in less used but open formats is absolutely compliant with the principle in discussion, as long as the only format used is not a proprietary one.

8. Data must be license-free:
   I.e. not subject to limitations deriving from IP rights, thus either in the public domain or disciplined by an open license.

Moreover, compliance with the Open Government Data principles needs to be demonstrable, i.e. there need to be accountability measures in place. The principles above are specific to the Open Government setting, and have therefore a higher degree of sector-specificity than the one of open data tout court. Their core, however, is the very same.

As resulting from the Open Definition, from the Open Government Data principles, and from the open data literature and documents available, some of which are discussed below, the

\[360\] For a side-by-side overview and explanation of both sets of principles, see Kitchin, The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences, 50 ss.

\[361\] E.g. the ‘G8 Open Data Charter’ (2013) \(<https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207772/Open_Data_Charter.pdf>\), according to which Open Data has to comply with five principles, titled as follows: Open Data by Default, Quality and Quantity, Usable by All, Releasing Data for Improved Governance, Releasing Data for Innovation. Another document that builds upon the 8 principles set during the Sebastopol meeting identifies the following 10 elements as benchmarks through which to evaluate openness and accessibility of open government data: completeness, primacy, timeliness, ease of physical and electronic access, machine readability, non-
main requirements that need to be fulfilled to qualify a dataset as open data, for the purposes of this thesis, can be summarized as being the following:

1. Unfettered availability, access, and usability;
2. No barriers to re-use, redistribution, and remixing;
3. Universal participation, equality, and non-discrimination;
4. Full machine readability, interoperability, and modifiability;
5. Based on open standards, formats, and licenses;
6. Free of charge (or close to it).

The notion of open data, as resulting from the elements above, highlights how the openness required, as it is often said, is twofold: on one hand, the data needs to be legally open, so either in the public domain or covered by an open license; on the other hand, the data needs to be technically open: in a machine readable and, when possible, non-proprietary format. In addition, recent open data literature and the access policy literature on which it builds highlight the importance of intellectual accessibility (the ability to understand the data), financial accessibility (the ability to afford the data), and ‘practical’ openness.

From the elements above, we can define open data as a set of requirements, or a data publication and re-use policy, intended at making the highest possible amount of data available to all and everybody, for any and every purpose or activity, in a machine-readable, interoperable, and modifiable format, based on open standards and digestible by free/libre open source software (FLOSS), and free of charge (or close to it). Any sort of restriction on the possibility to access, use, re-use, and distribute is not compatible with the requirements on which open data is based. Those requirements, which are supported through laws and policies all over the globe, embody an ideal – that public information should be free for all to use, re-use, and distribute – which is carried forth by a movement (i.e. the open data movement).

3.1.2 THE LEGAL STATUS OF OPEN DATA

Open data is thus information that is either in the public domain, or covered by an open license. In the former case, there is no right to (intellectual) property over the information to speak

\[\text{discrimination, use of commonly owned standards, licensing, permanence, and usage costs – John Wonderlich, ‘Ten Principles for Opening up Government Information’ (Sunlight Foundation 2010).}\]

\[\text{362} \text{ Wendy Carrara and others, ‘Open Data Goldbook for Data Managers and Data Holders’ (2016).}\]


\[\text{364} \text{ See Bastiaan van Loenen, Developing Geographic Information Infrastructures: The Role of Information Policies (IOS Press 2006) and the literature cited therein.}\]

\[\text{365} \text{ A detailed framework for assessing the success of open data initiatives, which builds upon and condenses several other assessment frameworks, has been provided in Frederika Welle Donker and Bastiaan van Loenen, ‘How to Assess the Success of the Open Data Ecosystem?’ (2017) 10 International Journal of Digital Earth 284.}\]

\[\text{366} \text{ E.g. copyright.}\]
of. In the latter, the information’s openness is ensured through a set of licensing terms (i.e. contractual conditions) through which the licensor removes all the limits to the possibility to access, use, and distribute the data that would otherwise be set by intellectual property law. In other words, when open data relies on licensing, the openness of the dataset derives from the employment of contractual mechanisms, and not from some derogation from intellectual property law. Those contractual mechanisms, or licensing conditions, render data ‘open’ by leveraging freedom of contract to create relative rights inter partes, and do not derogate from the applicable intellectual property regimes.

That is important to remember for two reasons. The first reason is that this thesis is concerned with exploring the idea of balancing open data and data protection, which is not a property-based regime like intellectual property.367 It is the rightsholder’s freedom of contract (i.e. private law) that allows an open data licensor to grant unfettered freedom to licensees by removing the constraints that would otherwise derive from intellectual property law. When data protection law (i.e. public law) applies, the information and its processing are subject to a number of additional constraints,368 such as for instance the purpose limitation principle,369 or the obligations relating to the security of the processing,370 that cannot be waived by data subjects like rightsholders license their intellectual property. The idea of balancing data protection and open data must thus move from the consideration that open data sharing and re-use does not derive from a deviation from intellectual property law, but rather from the use of licensing conditions and freedom of contract. Licensors, through freedom of contract, can grant licensees enough permissions vis-à-vis their intellectual property to allow the achievement of the requirements outlined in the previous subsection, so that the information considered is accessible, distributable, and reusable by anyone and for any purpose. The same lack of constraints cannot be achieved, as Chapters 4 and 5 will discuss, vis-à-vis data protection.

The second reason for which is important to keep in mind the normative mechanisms that make open data “open” in the sense discussed above derives from the fact that open data is often fostered by legislative means. A number of laws and regulations, in several jurisdictions all over the globe, oblige public bodies, and sometimes even private actors, to make available some of the information they collect during the performance of their main tasks. At the EU level, the main instrument371 promoting open data is the PSI Directive, which, since 2013, mandates that some categories of public-sector bodies372 must make available for re-use (some of) the documents they produce in their day-to-day operations. There is thus now, at

---

368 See Ch. 5.
369 GDPR, Art. 5(1)(b).
370 GDPR, Art. 5(1)(f); Art. 32.
371 Although other EU laws play a role in stimulating open data sharing and re-use (see the following sections), and Member States’ national legislation (e.g. freedom of information (FoI) laws) add an additional normative layer.
372 And, since its 2019 recast, some private-sector actors too – see below.
EU level, a ‘right to re-use’ PSI, which has been increasingly tied with the concept of open data. Such right is obviously bound to interact with all the other regulatory regimes that may apply to the processing of a given dataset – data protection, intellectual property, or competition law, for instance. While the EU PSI regime is instrumental to open data sharing and re-use, it is important to keep in mind that it is not PSI law that renders a given dataset “open data”, but (inter alia) its licensing regime, or the fact that it is in the public domain.

That aside, again, open data needs to be either in the public domain, or bound by an open license. Such a license, which should preferably be compatible with other kinds of open licenses, must allow the use and redistribution\(^{373}\) (on its own or as part of a collection) of the data, and its modification and subsequent redistribution, if under the same terms covering the original dataset. All the above applies both to the dataset in its entirety, and to parts of it, which means that the license must allow any part of the dataset to be freely used, distributed, or modified separately from any other dataset in which it was originally contained; conversely, the license must allow the dataset to be distributed within other datasets, without placing any restriction on them. Furthermore, the license must not discriminate, either at an individual level or at a collective one, thus treating all potential re-users equally, and must automatically grant the rights attached to the work to all the re-users, without the need for them to agree to any additional terms and conditions. The dataset must be made available for reuse for any purpose desired by the re-user, and thus no teleological or sectorial limitations are to be sanctioned by the license agreement. Finally, the license must not foresee any form of monetary compensation or remuneration for the reuse of the dataset.

There are, however, some conditions that can be attached to an open license without depriving it of its ‘open’ status. For starters, clauses requiring further distributions of the datasets to attribute its original contributors, rights holders, sponsors, and creators are deemed compatible with the concept of open data. The license can also foresee that further distributions indicate their status through a different name or number, or otherwise indicate that the original dataset has been changed or distributed, thus specifying the original’s integrity. Furthermore, the license can have a share-alike clause, binding all further distributions to be bound by the same kind of open license; to this purpose, it may also require the retention of copyright notices and the identification of the open license employed. Clauses binding further distributors to provide users with the preferred format for making modifications are allowed, and so are clauses forbidding the employment of technical protection measures excluding the exercise of rights otherwise granted under the open license. Finally, non-aggression clauses\(^{374}\) may be foreseen in open licenses as well. It seems interesting to note, in this respect, how one of the most contested points is whether, in the

\(^{373}\) Sale included.
\(^{374}\) “The license may also condition permissions on not aggressing against licensees with respect to exercising any allowed right (again, for example, patent litigation)” – ‘Open Definition 2.1’, par. 2.2.7. About non-aggression clauses, see Catharina Maracke and Axel Metzger, ‘Playing Nice with Patents: Do Voluntary Non-Aggression Pledges Provide a Sound Basis for Innovation’ (2015) 17 NCJL & Tech. 483.
spirit of Copyleft\textsuperscript{375} licensing, “share-alike” clauses,\textsuperscript{376} which require further distributions of the dataset to remain under the same license or a similar open license, are in accordance with the values underlying open data. While those clauses preserve the openness of a dataset’s derivative works, they also constrain its re-use for a number of purposes (e.g. commercial purposes), and restrictions by field of endeavour is incompatible with the concept of open data.

3.2 Open Data: Some Historical and Genealogical Notes

The brief review of what makes the concept of open data undertaken above has been performed taking the Open Definition as a starting point.\textsuperscript{377} As this section aims at briefly giving some contextual historical and genealogical\textsuperscript{378} notes, it appears appropriate to start by mentioning the roots of the Open Definition itself, as specified by its authors: “The Open Definition was initially derived from the Open Source Definition, which in turn was derived from the original Debian Free Software Guidelines, and the Debian Social Contract of which they are a part, which were created by Bruce Perens and the Debian Developers. Bruce later used the same text in creating the Open Source Definition. This definition is substantially derivative of those documents and retains their essential principles. Richard Stallman was the first to push the ideals of software freedom which we continue”.\textsuperscript{379}

The idea underlying open data’s openness derives from the open source movement,\textsuperscript{380} and from the instances it pushed (and pushes) forth, which results clearly from the references to FLOSS and to open formats and standards characterizing the current understanding of the concept of open data.\textsuperscript{381} There are, however, FLOSS movement aside, several other factors that contributed in shaping open data, both from an historical and from a genealogical perspective. Indeed, the same degree of merit in introducing and evolving the notion of open data as understood today must be given to the Open Access and Open Science movements, which aimed at rendering scientific output and governmental information – in particular if

\textsuperscript{375} Copyleft is the idea of using IP and contract law to grant the right to freely distribute a given work (from software to datasets to art) and its modifications, provided that the same conditions are kept in further distributions and derivative works; see e.g. Robert W Gomulkiewicz, ‘How Copyleft Uses License Rights to Succeed in the Open Source Software Revolution and the Implications for Article 2B Symposium: Licensing in the Digital Age’ Houston Law Review 179.

\textsuperscript{376} “Viral” clauses mandating further distributions and modifications of the work licensed to be bound by the same (open) license, or by a compatible one. For the same purpose, some open data definitions allow to prescribe through licensing that further distributions and modifications of the dataset remain free of any technical protection measures.

\textsuperscript{377} The Open Definition speaks broadly of ‘work’, rather than of data, defining it as “the item or piece of knowledge being transferred”. It still applies, \textit{mutatis mutandis}, to open data.


\textsuperscript{379} Attribution appended to Open Knowledge International (n 32).

\textsuperscript{380} The parallelism results clearly from the four fundamental freedoms of free software, as framed by Richard Stallman and the Free Software Foundation: “The freedom to run the program as you wish, for any purpose (freedom 0); The freedom to study how the program works, and change it so it does your computing as you wish (freedom 1); The freedom to redistribute copies so you can help your neighbor (freedom 2); The freedom to distribute copies of your modified versions to others (freedom 3)” – Free Software Foundation, ‘What Is Free Software?’ <https://www.gnu.org/philosophy/free-sw.html> accessed 23 June 2016.

\textsuperscript{381} See Open Knowledge International (n 32). See also Jeffrey Pomerantz and Robin Peek, ‘Fifty Shades of Open’ (2016) 21 First Monday.
publicly funded – as free from the limitations deriving from IP rights as reasonably possible, shaping scientific publication just as the FLOSS movement aimed at doing with software. Moreover, the open data movement has been developing paralleling the Right to Information (RTI) one, which is mainly concerned with enhancing transparency and accountability within public administrations and other public bodies and agencies, and whose connection with the notion of open data results therefore apparent. Aside from RTI instances, furthermore, the open data movement gained additional traction by leaning on the social, economic and political ratio underlying PSI reuse: the potential for increased efficiency and for added value and growth deriving from the reuse of PSI pushed open data instances within environments in which transparency and openness would not have had the same level of traction.

The last 20 years saw the open data movement growing from relatively narrow academic and professional circles to a full-fledged movement that is arguably changing how both the public and the private sectors operate, often in a radical fashion. The distributed, inclusive and non-hierarchical nature of open data instances, coupled with the fact that the concept is of relatively recent birth, renders the development of a linear timeline a somehow sterile exercise in comparison with a genealogical inquiry; however, there are still several specific events that were paramount in shaping the concept of open data and the movement underlying it, as understood nowadays. These events will be briefly accounted for in the following subsections.

### 3.2.1 Europe

The open data movement progressively gained a foothold in Europe as well: as highlighted in the European Commission’s communication “Open data: an engine for innovation, growth and transparent governance”, several national and European initiatives – both of a legislative

---

382 Kitchin, The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences, 48.
383 PSI is logically connected to, and yet distinct from, the notion of Open Data: see Carrara and others (n 362). The chapter elaborates on the similarities and the differences between the two concepts.
385 See Gray (n 378).
387 Arguably, the development of open data instances and of the open data movement in general within the EU owes as much to national experiences – e.g. the UK’s one – as it does to European-wide initiatives. Despite the fact that further comparative research about the evolution of the concept of open data (and of the one of PSI) within individual Member States could be beneficial in developing a historical and genealogical understanding of open data, it would be beyond the scope of this thesis, at this stage, to delve into it. Therefore, this section is concerned with the EU’s open data policy, and does not consider the contribution that individual Member States had towards the concept and towards the movement.

74
and of a non-legislative nature – contributed in shaping the notion of open data and its implementation this side of the pond, particularly in connection with the economic exploitation of PSI. As a matter of fact, the Commission’s efforts in opening PSI has been ongoing since the end of the ‘80s, and the whole European institutions apparatus got progressively involved with the process, supporting it with both legislative and non-legislative measures. From a regulatory perspective, the concerns about the underdevelopment of the European information market when compared to the US one drove the EU legislative bodies to enact a few Directives, most notably the PSI Directive, to foster reuse of information within the Union.

The PSI Directive, which is purposed to enable the availability of PSI to third parties and to ensure fair competition between the public sector and the private information industry, is the result of a gradual, slow, and convoluted process that had a sizable influence on the development of open data and PSI in general. As it has been reported, the process’s inception was the “Guidelines for improving the synergy between the public and private sectors in the information market” developed in 1989 by the Commission of the European Communities’ Directorate-General for Telecommunications, Information Industries and Innovation, which aimed at providing a basic set of principles and recommendations to be used in the development of national guidelines. Despite being a landmark in EC access policy, and the result of several years of studies and discussions with the EC’s information market players, the Synergy Guidelines were however highly criticised for a number of factors, most notably the fact that they were unbalanced towards the needs of the private sector, and failed to meet their declared objectives.

Accordingly, a few years later, the Commission funded a series of three reports – known as the “PUBLAW Reports” – and of a related workshop, meant to thoroughly investigate the policies and practices of individual Member States regarding the information market. The Reports highlighted a general lack of awareness regarding the Synergy Guidelines and a high level of difference in their implementation amongst Member States, despite the stakeholders’ positive reaction to their promulgation. The lack of general policies with regard to the information market was found to be widespread amongst the EC’s components, and thus the


390 Janssen and Dumortier (n 388).


393 Janssen and Dumortier (n 388).

need for a common framework was recognized. Moreover, the Reports criticized non-binding force of the Guidelines, highlighting the opportunity of using a legislative instrument – e.g. a Directive – to achieve the same goals.

Despite the lacking results deriving from the non-binding efficacy of the Synergy Guidelines, the Commission kept pursuing the same road – i.e. the utilization of non-binding instruments – by issuing a Green Paper on Public Sector Information in the Information Society,395 whose need was openly expressed during the Stockholm conference “Access to public information. A key to commercial growth and electronic democracy”, held by the Commission in June 1996. The Green Paper, released in 1998, was based on a two-year consultation effort, and highlighted a number of issues396 pertaining to the access to and the reuse of PSI, which would have needed to be solved in order to achieve a common European policy on PSI exploitation. The stakeholders’ engagement in the consultation was extensive, and the feedback the Green Paper received was overall positive.397

Following the comments received by the relevant stakeholders after the Green Paper’s publication, the Commission issued a Communication titled “eEurope 2002: Creating a EU Framework for the Exploitation of Public Sector Information”,398 embedded within the eEurope Action Plan 2002, which focused on the economic and internal market aspects of PSI. The communication recognized the impact of the advent of the so-called knowledge-based economy, and frames PSI as an asset for EU citizens, administration, and businesses; it also acknowledged that the high regulatory fragmentation the EU was, at the time, subject to, deeply undermined the development of pan-European information-based products and services. Therefore, as part of the action plan proposed, the Commission propositioned the adoption of a minimum-harmonization legislative399 instrument – preferably a Directive – to ensure a level playing field within the information market across the EU.

After the communication was issued, on the basis of a working draft circulated at the beginning of 2002, the Commission held another consultation concerning the proposal of a legal instrument regulating the reuse of PSI, which successively became the PSI Directive. The PSI Directive, its 2013 amendment, and the 2019 recast, will be discussed further in section 3.3 below, as its structure is best read in conjunction with its history and legislative development.

The development of the concept of open data within the EU has been shaped by its open data strategy as well. Contained in the Commission’s communication “Open data – An engine for innovation, growth and transparent governance”,400 along with its proposal for a revised PSI

396 E.g. IP rights, personal data protection, pricing, liability issues, exclusive licensing.
397 The 2000 ‘PIRA Report’ had also a significant role in gathering of the political willingness necessary to move forward with the introduction of pan-European PSI legislation: see Emma Broomby and others, ‘Commercial Exploitation of Europe’s Public Sector Information: Final Report’ (2000).
399 The Commission’s communication explicitly underlined the failure of the Synergy Guidelines, framing it as deriving from the non-binding nature of the instrument: “the experience with the 1989 Guidelines issued by the Commission services on the exploitation of public sector information suggests that a non-legislative approach in this area will not lead to the necessary results. In practice these guidelines have had little impact”.
400 European Commission, ‘Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and The Committee of The Regions “Open Data An Engine for Innovation, Growth and Transparent Governance” COM(2011) 882 Final’ (n 386).
Directive and with a revision of the Commission’s decision on the reuse of its own information\textsuperscript{401}, is part of the wider Digital Agenda for Europe program. The EU open data strategy is framed as consisting of three intertwined ‘strands’: a revision of the legal framework for data re-use, financial support for the development of data-portals and tools, and facilitation of coordination and sharing across the EU’s member states. It explicitly recognizes open data’s potential and the achievements reached by the EU and by its members in this respect, but also the obstacles and hurdles the development of open data and PSI reuse within the EU was still subject to. In particular, the strategy indicates the fragmentation of the legal framework, the general lack of awareness amongst public and private organisations, and a number of practical and technical issues\textsuperscript{402} as the main concerns remaining regarding open data developments within the EU.

As part of the EU’s open data strategy, non-legislative EU-wide measures have been enacted as well, and they contributed sensibly in developing the notions of PSI and of open data too. Those measures, which are manifold and whose extensive account is beyond the scope of this thesis, could be roughly catalogued as belonging to four different categories: research initiatives, training and education, development of portals and tools, and funding/economic stimulation. As for the first category – research initiatives – a prime example would be the LAPSI (Legal Aspects of Public Sector Information) and LAPSI 2.0 thematic networks, funded by the EU, which ran from January 2010 until December 2012 and from January 2013 until December 2014 respectively, and brought together scholars, practitioners, and representatives from the public sector. The project contributed significantly in framing and clarifying PSI and open data within the EU.\textsuperscript{403} Moreover, the EU has, over the past decade, funded a number of studies aiming at clarifying PSI and open data and their societal and economic significance,\textsuperscript{404} which mainly focused on the analysis of the economic impact of PSI and on the presence of exclusive conditions (which could hamper the development of a level playing field within the EU information market) within individual Member States.

The initiatives aiming at provision of training and education by entities connected to or embedded within the EU institutions arguably play and played a role in developing the notions of PSI and open data as well; amongst them, the (recently discontinued) ePSI Platform – an online platform that provided users with news, best practices, use-cases and legal

\textsuperscript{402} e.g. language barriers and interoperability.
\textsuperscript{403} For an overview of the LAPSI and LAPSI 2.0 thematic networks’ deliverables, see European Commission, ‘Legal Aspects of Public Sector Information (LAPSI) Thematic Network Outputs’ (2015).
developments regarding PSI – is worth a mention. Furthermore, an expert group on PSI (the PSI Group) was set up in 2002, and purposed at exchanging best practices and initiatives to foster information exploitation within the EU. Research aside, the EU has been involved in funding private ventures dealing with PSI and open data as well: for instance, through the open data Incubator for Europe (ODINE), a Horizon 2020-funded programme set up to finance an incubator supporting open data entrepreneurial activities in Europe.

Finally, the EU has been active in developing and deploying a number of tools – web portals prominently amongst them – to foster PSI reuse and open data from an operational perspective. Open data portals are web-based interfaces that allow the finding and exploitation of PSI. They constitute information systems (sometimes systems of systems) that, through the adoption of uniform metadata schemes (to catalogue and sift through resources efficiently and effectively) and of APIs (which allow to easily develop applications that automatically interact with the portal), thus either facilitating or allowing tout court the reuse of PSI. In particular, the creation of both the European Union Open Data Portal (a single point of access to data from the institutions and other bodies of the EU) and the European Data Portal (which harvests the metadata of PSI hosted in public data portals within EU Member States) has been paramount in the evolution of open data practices within the Union: openness is to be indeed constructed from a technical perspective as much as from a conceptual one, and portals are a major step forward in that sense.

3.2.2 INTERNATIONAL DEVELOPMENTS

Aside from the traction received from the policy and legislative developments in the United States and within the European Union, open data has been pushed by a multiplicity of actors at an international level as well.

---

405 The platform, originally reachable at http://www.epsiplatform.eu, has been discontinued at the beginning of 2016, its content moved to the European Data Portal. As the ePSI Platform’s notice went, “the ePSI Platform is no more, but the data lives on”.
408 On the topic, see Jeffrey Pomerantz, Metadata (MIT Press 2015). See also Vassilios Peristeras, ‘Towards Open Government Metadata’ (2011). The format used by the European Data and European Union Open Data portals is the DCAT Application profile for data portals in Europe, (DCAT-AP), which derives from DCAT, the Data Catalogue vocabulary.
411 Many European Member States have their own open data portals, some of which gather data at a national level, and some of which do it on a local level, e.g. the French opendata.paris.fr and www.data.gouv.fr; the Italian regional www.dati.piemonte.it and national www.dati.gov.it; the Dutch www.data.overheid.nl, and finally the UK’s data.gov.uk, historically a frontrunner in the EU’s open data development.
412 The importance of portals and tools to promote open data is confirmed, inter alia, by the PSI Directive’s recital 23: “Tools that help potential re-users to find documents available for re-use and the conditions for re-use can facilitate considerably the cross-border use of public sector documents. Member States should therefore ensure that practical arrangements are in place that help re-users in their search for documents available for re-use. Assets lists, accessible preferably online, of main documents (documents that are extensively re-used or that have the potential to be extensively re-used), and portal sites that are linked to decentralised assets lists are examples of such practical arrangements”.
The OECD “Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information”, developed in 2007 on the basis of prior work and adopted by the OECD Council in 2008, for instance, invites the OECD’s member countries to take account of a number of principles in establishing or reviewing their policies regarding PSI access and reuse. Those principles provide a general framework that, in accordance with the foundational elements of the concepts of open data, aim at fostering PSI reuse and access to government information. According to the Recommendation, member countries are invited to adopt the following principles:

- **Openness**, which substantiates in maximising the availability of PSI by adopting openness as the default setting for government data releases, wherever possible and no matter the model of funding is, to boost access and re-use.

- **Access and transparent conditions** for re-use, meant to foster broad, non-discriminatory, and competitive access to PSI, eliminating exclusive arrangements and unnecessary restrictions on access and reuse, “so that in principle all accessible information would be open to reuse by all”, preferably over the Internet and in electronic form.

- **Asset listing** to generate awareness, both of what public sector information is available e.g. through online lists and inventories, and of the conditions to access and re-use the available PSI.

- **Data quality and integrity**, through the development and adoption of data management best practices.

- **New technologies and long-term preservation**, to improve interoperable archiving, search and retrieval technologies and related research, addressing technological obsolescence and finding new ways for PSI digitisation.

- **Respect for IP rights**, for there are many ways to open up PSI, ranging from governments or private entities holding copyrights, to public sector information being copyright-free.

- **Transparent and uniform pricing**, in case PSI is not provided free of charge, possibly not exceeding the marginal costs of maintenance and distribution.

- **Fair and neutral competition**.

- **Provision of appropriate redress mechanisms**.

- **Facilitation of public-private partnerships** (PPPs).

- **Fostering international, cross-border access** and reuse of PSI.

- **Encouraging the development and sharing of industry-wide best practices**.

The Recommendation’s content development built on preceding exploratory studies and workshops organized by the OECD, such as a workshop on PSI specifically held to prepare the

---


414 The Recommendation of course recognizes the existence of legitimate grounds of refusal or limitations, e.g. the protection of national security interests, privacy and data protection, intellectual property, or other conflicting national laws.

Recommendation (2008); a study on public sector information and content; and a conference on the Future Digital Economy: Digital Content Creation, Distribution and Access, jointly organised by the OECD and the Italian Minister for Innovation and Technologies (2006).

The 39th Group of Eight (G8) summit, held in June 2013 at the Lough Erne in County Fermanagh, Northern Ireland, significantly contributed in the diffusion of open data as well. During the summit, the G8 leaders signed the G8 Open Data Charter, which contains a set of five core principles meant to be “the foundation for access to, and the release and re-use of, data made available by G8 governments.” The principles developed are the following:

1. **Open Data by Default**, unless specific reasons dictate otherwise.
2. **Quality and Quantity**, meaning that it is imperative to aim at releasing as much information as possible, and of the best quality, timeliness, and documentation achievable.
3. **Usable by All**, without discriminating neither the possible users of the information released nor the means through which those users could access and process that information.
4. **Releasing Data for Improved Governance**, fostering both transparency and efficiency within the public sector.
5. **Releasing Data for Innovation**, fostering economic growth and scientific progress.

The principles are to be implemented in accordance with the best practises and timeframes set out in the G8 members’ national action plans. The G8 Open Data Charter has also a technical annex, divided in two parts: an enumeration of best practices through which to substantiate each of the Charter’s principles, and a set of collective actions to be performed by the G8’s members. As for the latter part, the actions that the Charter mandated to its adherents are three: implementing the charter in a national action plan, releasing as many high-value datasets as possible, and contributing to and maintaining both the metadata mapping index across the G8 member states’ metadata and the detailed page on each state’s use of metadata.

---


420 Group of 8 (n 361).

421 Group of 8 (n 361).

422 As indicated in the G8 Open Data Charter’s Annex.


The G8 Charter had a generally positive reception, but the international open data community felt that, despite the right direction taken by the Charter, there was still room for specification and improvement. Therefore, in 2015, several international organisations and governments came together, and developed the International Open Data Charter.\textsuperscript{425} The works for the International Charter started in May 2015, at the Open Data Conference in Ottawa. The Open Government Partnership (OGP) Open Data Working Group,\textsuperscript{426} the governments of Canada and of Mexico, the Web Foundation, the International Development Research Centre, the Open Data for Development Network, and Omidyar Network convened a meeting which constituted a group of “stewards”, composed by representatives of governments and civil society organisations from around the world.\textsuperscript{427} At the meeting, in which over a thousand participants convened, an action plan\textsuperscript{428} for the International Charter’s development was agreed upon; subsequently, a two-month consultation of the draft document of the Charter was launched. The Charter, which was rolled out at a number of events\textsuperscript{429} at the end of 2015, constitutes of six principles for the release of open data, which shall be:

1. Open by Default;
2. Timely and Comprehensive;
3. Accessible and Useable;
4. Comparable and Interoperable;
5. For Improved Governance and Citizen Engagement; and
6. For Inclusive Development and Innovation.

The International Charter is open for adoption by national\textsuperscript{430} and local\textsuperscript{431} governments and for endorsement by non-governmental organisations. It is established as a “Global Multi-Stakeholder Action Network” led by “Stewards”\textsuperscript{432} and “Lead Stewards”,\textsuperscript{433} whose members

\textsuperscript{426} The OGP is an international multilateral initiative that aims at promoting transparency, civic participation, and innovation at the governmental level. The promotion of open data policies and practices is one of its core areas of action.
\textsuperscript{428} IODC, ‘Enabling the Data Revolution – An International Open Data Roadmap’ (2015).
\textsuperscript{429} In particular, the OGP Summit in Mexico City (October 27-29, 2015), the G20 Leaders’ Summit in Turkey (November 15-16, 2015), and the COP21 in France (December 7-8, 2015).
\textsuperscript{430} Argentina, Chile, France, Guatemala, Italy, Mexico, Panama, Philippines, South Korea, United Kingdom, Uruguay.
\textsuperscript{431} City of Buenos Aires, City of Madrid, Municipality of Minatitlán, Municipality of Puebla, Municipality of Reynosa, Municipality of San Isidro, Municipality of Torrón, Municipality of Xalapa, State of Mexico, City of Quito, Municipality of Veracruz, State of Morelos.
\textsuperscript{432} The governments of France, Italy and the city of Buenos Aires, the Sunlight Foundation, Avina, Open Alliance Nigeria, Hivos Hoofdkantoor, Gambia YMCAs Computer Training Centre & Digital Studio, YoungInnovations (Nepal), Caribbean Open Institute, The Governance Lab, Open Institute, The Center for Open Data Enterprise, the World Bank, United Nations, the Organization of American States (OAS), the Organization for Economic Co-operation and Development (OECD), the Inter-American Development Bank (IADB), the African Development Bank Group, the Economic Commission for Latin America and the Caribbean (ECLAC/CEPAL), and others.
are organised in Working Groups, and is coordinated, at the time of writing, by an Interim Secretariat.

Open data has been thus on the rise in both sides of the Atlantic, and lately also at an international level, as shown by the review above. The concept and the movement supporting it are reasonably well established across the globe, also due to the wide range of interests that open data aims at safeguarding: amongst others, transparency, participation, public sector efficiency, economic growth, open access. The following subsections will therefore disambiguate the concept of open data, initially moving from an introductory examination of the open data-related legislation in the EU and of the interests it aims at safeguarding.

3.3 Open Data Legislation in the EU

The sections above briefly outlined the recent developments of the concept of open data from a historical and genealogical perspective, taking into account the main policy developments that concerned the evolution of the open data movement. The concepts of open data and of PSI re-use are of course also highly dependent from their concrete legislative enactment; this section therefore examines how they are concretely implemented within the EU legislative milieu. As the PSI Directive’s evolution intertwines with the subsections above, and in consideration of its prominence amongst EU open data legislation, the following subsection will also give an account of its legislative trajectory.

3.3.1 The PSI Directive

As highlighted in the previous section, the development of open data within the EU – as a concept, as a licensing scheme, and as a movement – has been largely conditioned by the PSI Directive, both in its original version and as amended in 2013. What follows aims at clarifying the PSI Directive’s history and content, as it was in its original 2003 version, as amended in 2013, and as resulting from its 2019 recast as the new “on open data and the re-use of public sector information”.

3.3.1.1 History

The processes through which open data was institutionalized within the EU culminated in 2003 with the formal adoption of the PSI Directive, which, for the first time, regulated the European PSI market. The legislative process followed the ordinary legislative procedure. The Commission’s initial proposal was issued in June 2002, amended by both the European Parliament and by the Council of the European Union, agreed upon by the Commission, the Parliament, and the Council, and then finally signed into law on the 17th of November 2003.

In order to fulfil its purpose – enabling the economic potential deriving from the re-use of public information – the 2003 PSI Directive built on two pillars: transparency, on one hand, and fair competition, on the other; it aimed at harmonizing the conditions for PSI reuse across

---

436 Recital 25 identifies as objectives “to facilitate the creation of Community-wide information products and services based on public sector documents, to enhance an effective cross-border use of public sector documents
the EU, removing its major obstacles and providing uniform rules on key issues such as pricing, licensing, and exclusive arrangements. The 2003 PSI Directive should have been implemented by the EU’s member states in two years’ time, by the 1st of July 2005; however, not every member state, however, made the deadline in time, and it took a few more years and infringement procedures before the completion of the transposition of the PSI Directive within each state’s legal system.

The 2003 PSI Directive aimed at regulating an environment that is bound to change at an extremely fast pace, and in which data availability and technological progress grow at an exponential rate. It did not take long, therefore, for the PSI Directive to need a formal revision, also in consideration of the fact that – despite significant progress in the re-use of PSI within the EU’s internal market – significant barriers still existed. Accordingly, as sanctioned by the 2003 PSI Directive’s Article 13, the Commission carried out a review that ended in 2009, exploring the possibility to amend the Directive. However, the PSI Directive’s implementation was carried on by many EU member states with a considerable amount of delay: the prohibition of exclusive arrangements, for instance, only became fully applicable as from the very year in which the Commission’s review took place. Therefore, the Commission found that the PSI Directive had not yet had its full impact, did not perceive the necessity to amend the 2003 PSI Directive, and rescheduled a further review in 2012 at the latest, drawing attention on a number of critical areas that were still in need of further improvements: namely, the full and correct implementation and application of the Directive, exclusive arrangements,

by private companies for added-value information products and services and to limit distortions of competition on the Community market”; see also recitals no. 1, 5, 6, 15, 16.

See the first paragraph of the 2003 PSI Directive’s Art. 12: “Member States shall bring into force the laws, regulations and administrative provisions necessary to comply with this Directive by 1 July 2005. They shall forthwith inform the Commission thereof”.

Only four of them met the 2005 deadline; the Commission consequently opened 18 infringement cases, some of which resulted in the ECJ delivering a judgement for failure to implement the PSI Directive.

At the end of 2006, Austria, Belgium, Portugal, Spain and Luxembourg had still not implemented the PSI Directive in their national legislation; the Commission took therefore action against those countries, referring them to the European Court of Justice, which initiated infringement proceeding.

See Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information, Recital 5: “Since the first set of rules on re-use of public sector information was adopted in 2003, the amount of data in the world, including public data, has increased exponentially and new types of data are being generated and collected. In parallel, we are witnessing a continuous evolution in technologies for analysis, exploitation and processing of data. […] The rules adopted in 2003 no longer keep pace with these rapid changes and as a result the economic and social opportunities offered by re-use of public data risk being missed”.


“The Commission shall carry out a review of the application of this Directive before 1 July 2008 and shall communicate the results of this review, together with any proposals for modifications of the Directive, to the European Parliament and the Council”.


charging, fair competition, identification and availability of information resources, and conflict resolution mechanisms.

The second review of the PSI Directive took place in 2011, anticipating by one year the deadline set by the Commission in its 2009 Communication,\(^445\) in order to fit in with the Digital Agenda for Europe program. The 2011 review builds on a public consultation, which received a high amount of feedback; on several studies on the untapped value of PSI;\(^446\) and on an Impact Assessment,\(^447\) which focused on five critical areas: transparency, competition and pricing, the Directive’s scope, its implementation and application, and the redress mechanisms foreseen. The 2013 amendments to the PSI Directive, which originated from the Commission’s proposal\(^448\) for a Directive amending the PSI Directive and were enacted through Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013,\(^449\) are the result of the Commission’s 2011 second review. In 2014, the European Commission also issued a Commission notice titled “Guidelines on recommended standard licences, datasets and charging for the reuse of documents”.\(^450\) The guidelines are not binding, and aim at providing guidance on the best practices available within the three areas that were perceived as most relevant for the re-use of PSI in Europe after the 2013 reform – licensing conditions,\(^451\) datasets to be prioritised for release,\(^452\) and the parameters for marginal cost charging.

As mandated by Article 13 of the PSI Directive, the European Commission had to carry out a review of the Directive’s application before the 18\(^{th}\) of July 2018, communicating its results, along with any proposals for amendments to this Directive, to the European Parliament and to the Council. On the 25\(^{th}\) of April 2018, the Commission adopted a proposal to revise the PSI Directive for the second time in fifteen years. The Commission’s proposal for a recast of the PSI Directive was presented as part of a package of measures aiming to facilitate the creation of a common data space in the EU.\(^453\) The proposal resulted from an extensive public


\(^{446}\) e.g. Vickery (n 404).


\(^{451}\) Interestingly, the Guidelines explicitly refer to the Open Definition provided by Open Knowledge International (n 32). The increased convergence between PSI law and the open data movement has been noted before by Janssen and Hugelier (n 434).

\(^{452}\) Geospatial data, Earth observation and Environment, Space and in situ data, Transport data, Statistics, Companies data.

consultation process, an evaluation of the 2013 PSI Directive and an impact assessment. The Commission’s proposal for a recast of the PSI Directive followed the ordinary legislative procedure and was amended by the European Parliament and by the Council, eventually becoming the 2019 Directive on Open Data and Public Sector Information.

3.3.1.2 STRUCTURE

The 2003 PSI Directive, which followed a relatively unobtrusive minimum harmonization strategy, consisted of 15 articles that aim at facilitating the creation of an EU-wide market for information products and services based on PSI by “establishing a minimum set of rules governing the re-use and the practical means of facilitating re-use”. It applied to every document held by public sector bodies, meaning that any content, and any part of such content, whatever its medium, are enclosed within the scope of the PSI Directive. However, the 2003 PSI Directive explicitly excludes from its scope the documents the supply of which falls outside the scope of the public task of the public sector bodies concerned; the documents for which third parties hold intellectual property rights; the ones that are excluded from access by virtue of the access regimes within individual Member States, e.g. on the grounds of national security, defence, or public security, or statistical or commercial confidentiality; documents held for the fulfilment of a public service broadcasting remit; documents held by educational and research establishments, and documents held by cultural establishments. The 2013 amendments refined the scope of the Directive, extending it to documents held by libraries (including university libraries), museums, and archives, but explicitly excluding parts of documents containing only logos, crests, and insignia, and specifying the exemption covering the documents that cannot be accessed or re-used on the grounds of protection of personal data.

---

460 See Art. 1(1) of the 2003 PSI Directive.
461 See Art. 3(3) of the 2003 PSI Directive.
462 See Art. 1(2) of the 2003 PSI Directive.
463 The 2003 PSI Directive, in its Article 1.2 a), conceived the public task “as defined by law or by other binding rules in the Member State, or in the absence of such rules as defined in line with common administrative practice in the Member State in question”; the 2013 amendments specify that common administrative practices can define the public task “provided that the scope of the public tasks is transparent and subject to review”.
464 See Art. 1(3) of the 2003 PSI Directive.
Article 3 of the PSI Directive codifies the Directive’s main, general principle – the existence of a right to re-use PSI. Originally, Member States were to ensure, *where the re-use of documents was allowed*, that these documents were re-usable for both commercial or non-commercial purposes and, where possible, made available through electronic means. The 2013 amendments however sanctioned the existence of a true ‘right to re-use’, mandating that all the information not explicitly covered by one of the exemptions mentioned above is to be made available by the public administrations concerned. An exception was introduced for documents in which libraries, including university libraries, museums and archives hold intellectual property rights, which are still subject to the pre-existing regime (i.e. Member States are still allowed to decide which of those documents are to be made available for re-use). Nor the original 2003 PSI Directive, nor the 2013 amended version, however, foresee the duty for Member States to proactively make available PSI in the absence of a specific request, since the PSI release depends on an explicit request by a potential re-user.

Public sector bodies must make their documents available in any pre-existing format or language, and, where possible and appropriate, in an open and machine-readable format, together with the metadata describing them. Both the format and the metadata should, if possible, be based on open standards. This does not imply, however, any obligation for public sector bodies to create or adapt documents or provide extracts in order to comply, in case it would involve disproportionate effort, going beyond a simple operation. It does not mean that public sector bodies can be required to continue the production and storage of a certain type of documents, either, which is explicitly beyond the scope of the Directive and of the duties it imposes on public administrations.

Public sector bodies are allowed to charge PSI re-users for the information released. The 2003 PSI Directive allowed charges not exceeding the cost of collection, production, reproduction and dissemination, together with a reasonable return on investment. The upper limit was lowered by the 2013 amendments, which adopted the marginal cost parameter, thus limited to the expenses incurred for PSI reproduction, provision and dissemination. The provision mandating the adoption of the marginal cost pricing limit does however not apply to public sector bodies that are required to generate revenue to cover a substantial part of their costs relating to the performance of their public tasks; to documents for which the public sector body concerned is required to generate sufficient revenue to cover a substantial part of their costs, as defined by law, by other binding rules or by common administrative practice in the Member State; and to libraries (including university libraries), museums and archives. Article 8 of the PSI Directive, furthermore, allows public sector bodies to impose conditions, e.g. through a license, to re-users. If licenses are used, Member States shall ensure that they are available in digital format, and encourage all public sector bodies to use standard licences.

Transparency must be ensured, with regard to the calculations of both standard and non-standard charges, and to the information regarding the means of redress available for re-users.

---

466 See Art. 1(3) of the 2013 Directive amending the PSI Directive.
467 See Art. 4 of the PSI Directive.
468 See Art. 5 of the PSI Directive.
469 See Art. 5(1) and 5(2) of the PSI Directive.
470 See Art. 6 of the 2003 PSI Directive.
472 See Art. 8(2) of the PSI Directive.
473 See Art. 7 of the PSI Directive.
that see their request denied. Moreover, the conditions for the re-use of similar categories of data have to be non-discriminatory, and if the documents are re-used by a public sector body for commercial activities that fall outside the scope of its public tasks, “the same charges and other conditions shall apply to the supply of the documents for those activities as apply to other users”.\footnote{See Art. 10 of the PSI Directive.} Furthermore, exclusive arrangements are prohibited:\footnote{See Art. 11 of the PSI Directive.} re-use is open for all market players, and the contracts stipulated by public bodies with third parties cannot grant exclusive rights concerning the reuse of PSI. If, however, an exclusive right is necessary for the provision of a service in the public interest, such a right can be granted, provided that the validity of the reason for granting such an exclusive right is subject to regular review, at least every three years. Exclusive right relating to the digitisation of cultural resources have a partly different regime: their duration should in general not exceed 10 years; when it does, its duration has to be reviewed during the 11th year and every 7 years thereafter. In any case, the exclusive arrangements established are to be transparent, and thus made public. In the specific case of the concession of an exclusive right covering the digitalization of a public resource, the public sector body concerned has to be provided with a free copy of the digitised resources as part of those arrangements, which will be available for re-use at the end of the period of exclusivity.

As sanctioned by Article 9 of the PSI Directive, finally, Member States are tasked with facilitating the search for documents available for re-use, e.g. by creating asset lists of main documents with relevant metadata, accessible where possible and appropriate online and in machine-readable format; they are also required to create web portals to access such asset lists, where possible facilitating the cross-linguistic search for documents.

The PSI Directive, whose history and structure have been sketched above, is one of the main instruments through which open data and the culture of PSI re-use have been advanced within the EU. What has been described in this subsection, on one hand, allows to highlight the fact that open data is as much dependant on policy instruments as much as it is on technological measures, and that within that context the two are bound to be intertwined, as the PSI Directive aptly exemplifies; on the other hand, it provides evidence of the multi-faceted nature of the concept of open data as conceived in EU policy, in that it explicitly focuses on the economical aspect of PSI reuse, rather than on transparency and on the stimulation of public sector efficiency, as this chapter will clarify later on.

3.3.1.3 The Recast of the PSI Directive

which it announced its proposal for a recast of the PSI Directive. The Commission’s proposal took into consideration an extensive amount of input: aside from the obligation ex Art. 13(2) of the 2013 PSI Directive, which mandated that Member States must submit periodic reports to the Commission on the implementation of the PSI Directive, the Commission engaged with stakeholders by launching a public consultation that spanned from June 2017 to late January 2018.

While the Commission found that the PSI Directive had been meeting its policy objectives of stimulating the digital single market and the cross-border exploitation of PSI while preventing distortions of competition, it also identified some areas that would need to be improved or modernised. Technological progress, legislative shortcomings in the 2013 amendments, and persisting normative fragmentation within the EU, rendered the PSI Directive ripe for change once again. The Commission’s preparatory works identified four main drivers, at the basis of an equal number of problem spaces, that have been tackled in the recast: insufficient access and re-use of dynamic data through APIs; market entry barriers and fragmentation; insufficient availability of public and publicly-funded data for re-use; and distortion of competition in the internal market. The recast, additionally, introduced a number of minor clarifications about the interaction between the PSI Directive and other legislative instruments, such as the Database, INSPIRE, and Public Access to Environmental Information Directives and the GDPR.

Each of the areas of intervention identified could have been tackled through measures having two different levels of intensity: a low-intensity packaged solution of regulatory interventions and soft law, or a high-intensity one. The Commission’s evaluation of the 2013 PSI Directive, the Impact Assessment, and the consultation process led to the choice of pursuing low-intensity, incremental change. While the high-intensity packaged solution explored by the Commission would have achieved similar results, its social and environmental impact would

480 The consultation process involved gathering feedback on the Inception Impact Assessment, a public online consultation, an extensive number of meetings, workshops, and roundtables, and a public hearing.
have been much higher, due to the greater amount of information to be subject to the PSI Directive and (therefore) the intensity of the obligations the latter would set.\footnote{For example, the high intensity option would have entirely eliminated the exceptions to the marginal cost pricing criterion, and included in the scope of the provisions regarding data in the transport and utilities sectors all economic operators which have been awarded a concession, rather than only public undertakings.}

**Scope extension: data in the transport and utilities sector and research data**

The legislature considered that, while EU Member States routinely entrust the provision of public services to private sector entities, the 2013 PSI Directive applied only to documents held by public sector bodies, and excluded public undertakings from its scope. This led to reduced availability of documents produced during the performance of services in the general interest, lowering PSI re-use’s potential in engendering innovation and growth.\footnote{Cfr. PSI Directive (recast), Recital 24.} The recast of the PSI Directive extends the Directive’s scope in a considerable manner by including in it public undertakings operating in the transport and utilities sectors. It does so through regulatory mechanisms similar to the ones adopted by the 2013 reform in respect to cultural institutions, which have no obligation to release the data they do not want to make available.\footnote{PSI Directive (recast), Art. 3(2); Recital 26.}

Following the low intensity approach chosen by the Commission, the recast covers only particular kinds of public undertaking.\footnote{I.e. “any undertaking active in the areas set out […] over which the public sector bodies may exercise directly or indirectly a dominant influence by virtue of their ownership of it, their financial participation therein, or the rules which govern it”; there is a presumption of dominant influence if public sector bodies, directly or indirectly, hold the majority of the undertaking’s subscribed capital, control the majority of the votes attaching to its shares, or can appoint more than half of the undertaking’s administrative, management or supervisory body – PSI Directive (recast), Art. 2(3).} Specifically,\footnote{PSI Directive (recast), Art. 1(1)(b).} only public undertakings active in the water, energy, transport and postal services sectors,\footnote{Directive 2014/25/EU of the European Parliament and of the Council of 26 February 2014 on procurement by entities operating in the water, energy, transport and postal services sectors and repealing Directive 2004/17/EC, 28.3.2014, OJ L 94/243.} acting as public service transport operators by rail or road\footnote{Pursuant to Art. 2 of Regulation (EC) No 1370/2007 of the European Parliament and of the Council of 23 October 2007 on public passenger transport services by rail and by road and repealing Council Regulations (EEC) Nos 1191/69 and 1107/70. 3.12.2007, OJ L 315/1.} or as air carriers fulfilling public service obligations,\footnote{Pursuant to Art. 16 of Regulation (EC) No 1008/2008 of the European Parliament and of the Council of 24 September 2008 on common rules for the operation of air services in the Community (Recast). 31.10.2008, OJ L 293/3.} and as Community ship owners fulfilling public service obligations.\footnote{Pursuant to Article 4 of Council Regulation (EEC) No 3577/92 of 7 December 1992 applying the principle of freedom to provide services to maritime transport within Member States (maritime cabotage). 12.12.1992, OJ L 364/7.} Documents held by public undertakings that are produced outside the scope of the general interest services they provide, or that are related to activities directly exposed to competition and therefore pursuant to Article 34 of Directive 2014/25/EU not subject to procurement rules, are however excluded from the new PSI Directive’s scope.\footnote{PSI Directive (recast), Art. 1(2)(b).} A more limited set of obligations is applicable to public undertakings, as compared to the general PSI regime. Public undertakings, for instance, can...
charge above marginal costs for dissemination, and are exempted from the general procedural rules on how to process requests for re-use.

The material scope of the PSI Directive has also been extended to research data, defined as any document in digital form, other than scientific publications, collected or produced in the course of scientific research activities and used as evidence in the research process, or commonly accepted as necessary to validate research findings and results. The PSI Directive however only applies to research that is publicly funded, and that has already been made publicly available through an institutional or subject-based repository. The recast is thus only applicable to research data for which the researcher has already made all relevant efforts to make the data publicly accessible, in particular through online databases. The previous exemption covering documents held by educational and research establishments, including organisations established for the transfer of research results, schools and universities, except university libraries, has been thus limited accordingly.

The re-use of research data is meant to be free of charge for the re-user, but educational establishments, research performing organisations and research funding organisations are also exempted from the general procedural rules on how to process requests for re-use. The new PSI Directive does not aim at the creation of a homogeneous set of rules on access and re-use of scientific information, leaving the implementation of the most appropriate open access policies to Member States’ discretion. Member States are to do so by following the ‘open by default’ and the FAIR principles, taking into consideration relevant concerns relating to intellectual property, data protection, security, and business confidentiality.

Dynamic data, APIs, and high-value datasets

Changes in the state of the art in data processing technologies led to amending the original PSI Directive in 2013 to foster open formats and standards, and to ensure technical interoperability and reusability. On the same line, the 2019 recast introduces an obligation for Member States to make dynamic data (“documents in a digital form, subject to frequent or real-time updates, in particular because of their volatility or rapid obsolescence”) available through suitable APIs, and where relevant as bulk download, immediately after collection.

---

497 PSI Directive (recast), Art. 6(2)(c).
498 PSI Directive (recast), Art. 4(6)(a).
499 PSI Directive (recast), Art. 1(1)(c); Recital 27-28.
500 PSI Directive (recast), Art. 2(9).
501 PSI Directive (recast), Art. 10(2).
502 PSI Directive (recast), Art. 6(6)(b).
503 PSI Directive (recast), Art. 4(6)(b).
504 According to which research data must be Findable, Accessible, Interoperable and Re-usable — see PSI Directive (recast), Recital 27. See generally Mark D Wilkinson and others, ‘The FAIR Guiding Principles for Scientific Data Management and Stewardship’ (2016) 3 Scientific data.
505 PSI Directive (recast), Art. 10(1).
507 PSI Directive (recast), Art. 2(8).
508 PSI Directive (recast), Art. 5(5). It may be argued that, rather than introducing a brand-new obligation, the recast of the PSI Directive just clarifies that these data were always meant to be under the scope of the PSI directive. With the pre-existing regime, it was possible to argue e.g. that dynamic data are generated after the re-user’s request is made, and do not exist at the time of the request, and one cannot request documents that do not exist at the time of the request. Another argument that could have been made in the pre-recast regime is that the definition of document in the 2013 PSI Directive implied that the data requested should be stored on
Proactive publishing of dynamic data is expected to result in a lessening of the administrative burden for public sector bodies, due to a lower number of re-use requests to process and a lower risk of complaints and litigation.\textsuperscript{509} Yet, public sector bodies can still be allowed to derogate from the obligations dynamic data provision immediately after collection where that would exceed their financial and technical capacity, and impose disproportionate effort, in which case the information will still have to be published in a timeframe that does not unduly impair its economic and social exploitation.\textsuperscript{510}

Some categories of dataset are associated with important socio-economic benefits that have a particular high value for society, the environment, and the economy, in particular because of their suitability for the creation of value-added services, applications, and jobs, and of the number of potential beneficiaries.\textsuperscript{511} For these datasets the recast of the PSI Directive introduces the category of ‘high-value’ datasets, whose reuse must be free of charge,\textsuperscript{512} and that must be made available by public bodies and public undertakings alike, in a machine-readable format, via suitable APIs and, where relevant, as a bulk download.\textsuperscript{513}

The categories of high-value datasets are set out in The PSI Directive’s Annex I,\textsuperscript{514} and are comprised of 6 themes: geospatial data, earth observation and environmental data, meteorological data, statistics, companies and company ownership information, and mobility data. The Commission is allowed to adopt, through a delegated act and under a narrow set of procedural conditions,\textsuperscript{515} a list of specific high-value datasets, together with the modalities of their publication and re-use. The Commission can, through such delegated act, specify the arrangements and conditions for the publication and re-use of high-value datasets, compatibly with open standard licences.\textsuperscript{516} The Commission’s implementing acts may also exempt specific high-value datasets held by public undertakings, where that would lead to a distortion of competition in the relevant markets.\textsuperscript{517} Member States may also exempt public sector bodies


\textsuperscript{510} PSI Directive (recast), Art. 5(6).

\textsuperscript{511} PSI Directive (recast), Art. 2(10).

\textsuperscript{512} PSI Directive (recast), Art. 6(6)(a).

\textsuperscript{513} PSI Directive (recast), Art. 5(8).

\textsuperscript{514} PSI Directive (recast), Art. 13(1).

\textsuperscript{515} PSI Directive (recast), Art. 14(1). The power to adopt delegated acts is conferred on the Commission for five years, starting from 16 July 2019, after which the delegation of power is tacitly extended for five more years, bar the opposition of the European Parliament or the Council. Not later than nine months before the end of the five-year period, the Commission must draw up an explanatory report. The delegation of power can be revoked, at any time, by either the European Parliament or by the Council. Such a decision terminates the power delegation from the day following its publication in the Official Journal of the European Union, or at a later date, if so specified, but without prejudice to the validity of any delegated act already in force. Before adopting a delegated act, the Commission must consult experts designated by each Member State. As soon as the delegated act is adopted, the Commission must notify the Parliament and the Council, who can object within three months starting from the notification of that act (a period that can be extended by three more months at the initiative of the European Parliament or of the Council). If there is no objection within that period, or if both the European Parliament and the Council have informed the Commission that they will not object, the delegated act enters into force.

\textsuperscript{516} PSI Directive (recast), Art. 14(1).

\textsuperscript{517} PSI Directive (recast), Art. 14(3).
that are required to generate revenue to cover a substantial part of the costs relating to their public tasks, where making the high-value data available free of charge would lead to a substantial impact on their budget, but only in the two years after the entry into force of the Commission’s implementing act.\footnote{518 PSI Directive (recast), Art. 14(5).}

**Charging and exclusive agreements**

The recast of the PSI Directive makes explicit the principle according to which the reuse of documents is meant to be free of charge.\footnote{519 PSI Directive (recast), Art. 6(1).} Accordingly, it narrows the exceptions to the principle stating that public sector bodies cannot charge more than the marginal costs for dissemination by removing the exemption covering documents for which the public sector body concerned is required to generate sufficient revenue to cover a substantial part of the costs relating to their collection, production, reproduction, and dissemination.\footnote{520 I.e. Art. 6(2)(b) of the 2013 PSI Directive.} It however also allows Member States to permit public bodies to charge for the recovery of the marginal cost incurred for the reproduction, provision and dissemination of documents, and the inclusion of the costs incurred to anonymize the dataset and to protect commercial confidentiality in the calculation of the marginal costs, thus partly relaxing the rules as well. Anonymisation\footnote{521 For an understanding of anonymization and anonymity in EU data protection law, see Article 29 Data Protection Working Party, ‘Opinion 05/2014 on Anonymisation Techniques WP216’ (2014); Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40); Article 29 Data Protection Working Party, ‘Opinion 06/2013 on Open Data and Public Sector Information (‘PSI’) Reuse WP207’ (2013).} is reckoned as the prime means to reconcile the interests underlying PSI reuse with the obligations deriving from EU data protection legislation, but the recast acknowledges how it might come at a cost.\footnote{522 PSI Directive (recast), Recital 52.}

The transparency requirements set by the PSI Directive have also been, to a certain extent, enhanced by its recast. Another novelty is Art. 12(4), which complements the norms applicable to exclusive agreements by mandating that any arrangement, either legal or practical, that aims at (or could reasonably lead to) restricting the availability of PSI even without expressly granting an exclusive right, must be published online at least two months before coming into effect, and be subject to regular reviews, in any event every three years.\footnote{523 PSI Directive (recast), Art. 12(4).}

The recast of the PSI Directive also foresees the \textit{ex ante} publication, online, of the list of the public sector bodies required to generate revenue to cover a substantial part of the costs relating to the performance of their public tasks,\footnote{524 PSI Directive (recast), Art. 6(2)(a)} as those bodies can derogate from the general charging regime.\footnote{525 PSI Directive (recast), Art. 6(3).}

**Clarifications on the interaction with overlapping legislation**

The recast of the PSI Directive also aims at providing explicit clarifications about the interaction between the PSI Directive and potentially overlapping legislation, such as the Database, INSPIRE, and Access to Environmental Information Directives, and the GDPR. While those clarifications do not actually change anything about the interaction between the PSI Directive...
and the other EU laws interested by the recast, they do provide additional guidance on how to solve possible conflicts between laws and on how to manage overlapping scopes.

The proposal clarifies the relationship between the PSI Directive and the sui generis database right ex Art. 7 of the Database Directive by specifying that such right should not be exercised by public sector bodies in order to prevent or restrict PSI re-use. It does not however alter the level of protection granted by the Database Directive to public sector bodies, nor changes the legal regime under the current 2013 PSI Directive – it merely clarifies how reliance on the sui generis database right should be avoided by virtue of how its protective rationale conflicts with the openness underlying the PSI Directive.

As for the INSPIRE Directive, which establishes a legal and technical interoperability framework for the sharing of spatial data held by public sector bodies, Art. 1(7) of the recast specifies how the PSI Directive is meant to regulate the re-use of all existing documents held by public sector, thus including documents to which the INSPIRE Directive applies. Consequently, geospatial data may fall under both the PSI Directive and the INSPIRE Directive. While INSPIRE focuses on access to services, data interoperability, and sharing of spatial information between administrations, the PSI Directive regulates the re-use of datasets, including geospatial data, so their overlapping application does not actually create major conflicts. Rather, as Recital 34 suggests, the requirements set for spatial information by the INSPIRE Directive can provide public sector bodies with a blueprint they can follow to ensure the compatibility, interoperability, and (hence) re-usability of the documents made available for re-use under the PSI Directive.

Recital 4 of the new PSI Directive, which enumerates the areas on which the substantive changes introduced by the recast focus, lists also the Access to Environmental Information (AEI) Directive amongst the regulatory instruments that may overlap with the PSI regime. However, the AEI Directive is mentioned just en passant: Recital 18 merely cites it as an example of the access regimes that are assumed to be consistent with the minimum harmonization operated by the PSI Directive, and to which the PSI Directive is not meant to be prejudicial.

The recast, finally, assumes that the interaction between the right to personal data protection and PSI re-use is already an established part of EU law, and does not foresee any major

526 PSI Directive (recast), Art. 1(6). See also Recital 61.
527 It has indeed been noted how INSPIRE, by complementing the general PSI regime, also fosters the development of open data within the EU – see Bastiaan van Loenen and Michel Grothe, ‘INSPIRE Empowers Re-Use of Public Sector Information’ (2014) 9 International Journal of Spatial Data Infrastructures Research 86.
528 Art. 2 of the INSPIRE Directive contains a non-prejudice clause directed both towards the PSI Directive (Art. 2(1)) and towards public authorities’ intellectual property rights (Art. 2(2)). In theory, INSPIRE and the PSI Directive have a different purpose and scope: if a public body makes spatial data available to other public institutions for its public task, the INSPIRE framework is applicable. If the public institution supplies spatial data to the private sector for re-use, or to another public institution for any use outside of its public task, the PSI directive applies. See Katleen Janssen, ‘INSPIRE and the PSI Directive: Public Task versus Commercial Activities’, Proceedings of 11th EC-GI & GIS Workshop, ESDI: Setting the Framework, Alghero, Sardinia (2005).
530 Notably, the Commission’s proposal did not – the meagre reference to the AEI Directive resulted from the following steps of the ordinary procedure.
addition in this respect. The new PSI Directive still specifies that it does not affect individuals’ right to personal data protection, and is without prejudice to EU and Member State law on the matter. Its recitals exemplify a potential overlap by declaring that the re-use of personal data is permissible only if the principle of purpose limitation is complied with, and warn that a data protection impact assessment Art. 35 of the GDPR may have to be performed before taking decisions about the re-use of documents containing personal data. Art. 1(4), a staple since the first version of the PSI Directive, still specifies how the PSI Directive was never meant to be prejudicial to EU and national data protection law.

The fact that the PSI regime is not meant to be prejudicial to data protection rules is also substantiated through Art. 1(2)(h), which excludes from the PSI Directive’s scope both documents access to which is excluded or restricted, by virtue of the applicable access regimes, on data protection grounds, and parts of documents that are accessible by virtue of those access regimes which contain personal data the re-use of which has been defined, by law, as incompatible with data protection legislation, or as undermining the privacy or the integrity of the individual. Compared with the 2013 PSI Directive, the new Art. 1(2)(h) merely introduces the following subparagraph: “[...] or as undermining the protection of privacy and the integrity of the individual, in particular in accordance with Union or national law regarding the protection of personal data”.

3.3.1.4 INCENTIVES AND PAST BEHAVIOUR: A PEEK AT THE EU DATA ECONOMY

The emergence of PSI policy and legislation within the EU has been a long and tortuous path, defined by the persistence of the EU Commission in trying to regulate the EU PSI market and by the expanding scope of its intervention. Moving from the non-binding Synergy Guidelines, the Commission reached a consensus on a binding instrument, the 2003 PSI Directive, and then in 2013 updated its general principle by introducing a right to re-use and extended its scope to cultural institutions. In 2019, the PSI Directive further increases its scope to public undertakings. Incentives and past behaviour are very good predictors for thing to come; this section thus discusses the recent Commission activity involving private sector data, and concludes with an educated prediction on the future direction of PSI regulation in the EU. Specifically, this chapter holds that the direction taken by the Commission reflects its aim of fostering access to private sector data through legislative means, and that the study of its past actions in the PSI domain might shed a light, mutatis mutandis, on the future of the regulation of private sector data sharing.

The 2019 proposal for a recast of the PSI Directive is part of a broader range of EU policy initiatives, and should be read in such context. The scope of the new PSI Directive is extended to certain kinds of public undertaking, in particular, which can be tied to the Commission’s work in the area of private sector machine-generated non-personal data, and to the 2017 Communication ‘Building a European Data Economy’. In that Communication, the

532 PSI Directive (recast), Art. 1(4).
534 PSI Directive (recast), Recital 52.
535 PSI Directive (recast), Recital 53.
536 European Commission, ‘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” (SWD(2017) 2 Final)’ (n 453).
Commission notes that an ever-increasing amount of information is generated by machines and processes based on emerging technologies, which is increasingly re-used to build new services, products, and processes, and to support data-driven decision-making. This presents lots of potential opportunities for operators in the EU data market. However, data is generally kept somewhat closed, processed in-house and often inaccessible to third parties. Such information is consequently analysed ‘in silos’, without the possibility to cross-correlate it with third party data. Access and reuse of machine-generated non-personal data is, for the Commission, thus central to the emergence of the EU data economy.

The Commission then engaged with the relevant stakeholders to explore possible actions on the matter. The Commission’s goals were to improve access to anonymous machine-generated data, facilitating and incentivising its sharing while at the same time protecting private companies’ investments and assets, minimising lock-in effects, and avoiding the disclosure of confidential or personal data.\textsuperscript{537} The regulatory solutions surveyed in the stakeholder dialogue were manifold and varied by virtue of their intensity: the Commission explored soft solutions like guidance on incentivising businesses to share data, the development of technical solutions, and default contract rules. It also explored harder regulatory options like mandating access for public interest and scientific purposes, creating ex novo a new ‘data producer’s right’, or implementing a framework granting access against remuneration similar to fair, reasonable and non-discriminatory (FRAND) terms in IP licensing.\textsuperscript{538} The outcomes of the stakeholder engagement highlighted\textsuperscript{539} generally mixed feelings about the options investigated by the Commission, and a cautious stance towards possible regulatory solutions. Most stakeholders agreed with the Commission’s objective of making more data available for reuse, but were very prudent in endorsing horizontal regulatory solutions in the field of private-sector data sharing.

The stakeholders’ feedback was clearly taken into account in the Commission’s following actions, namely the development of a series of measures, introduced by the 2018 Communication ‘Towards a Common European Data Space’,\textsuperscript{540} aiming at enabling the development of new products and services based on data. Aside from the recast of the PSI Directive, the measures issued included an update of the Recommendation on access to and preservation of scientific information\textsuperscript{541} and, more importantly, guidance on sharing private sector data in both a B2B and B2G context.\textsuperscript{542} The Commission’s interest in private sector data sharing substantiated itself through (proposed) binding regulation, albeit to a somewhat narrow extent, through the inclusion of public undertaking within the scope of the PSI Directive. The Commission, moreover, still pushed through with its aim of regulating private-

\textsuperscript{537} European Commission, ‘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” (SWD(2017) 2 Final)’ (n 453) 11.

\textsuperscript{538} European Commission, ‘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” (SWD(2017) 2 Final)’ (n 453) 12.

\textsuperscript{539} European Commission, ‘Synopsis Report Consultation on the “Building a European Data Economy” Initiative’ (2017).

\textsuperscript{540} European Commission, ‘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 European Data Space COM(2018) 232 Final’ (n 453).


sector data sharing, but took stakeholders’ resistance into consideration, and opted for issuing non-binding guidelines.

There is, arguably, a leitmotiv underlying the Commission’s interest in information exploitation, and it can be argued that its efforts in fostering private sector information sharing seem to mirror its previous activity in the public sector information area. As private-sector information sharing has been dealt with through the implementation of non-binding guidelines, PSI were initially regulated through the non-binding Synergy Guidelines. The stakeholders’ opposition to binding PSI regulation gradually waned away, leaving room for the drafting of the 2003 PSI Directive, and subsequently for its 2013 version, which introduced an actual ‘right to re-use’.

The Commission’s goal is still the same: fostering data sharing to further the single market and to avoid regulatory fragmentation. What changes is the target of the Commission’s intervention: public sector information back when the public sector was the main data collector and holder, and private sector data now that private companies are outshining public sector bodies in their data collecting and processing capabilities. Again, incentives and past behaviour are prime predictors for the future. The Commission has currently the same incentives for regulating private-sector data sharing that it had at the beginning of the history of EU PSI regulation, and the development of its Guidelines on private-sector B2B and B2G data sharing seems to mimic the development of the Synergy Guidelines in 1989. The development of PSI regulation can thus, perhaps, provide a blueprint of what is to come for private-sector data sharing, foreshadowing future regulation on private-sector information.

3.3.2 THE INSPIRE DIRECTIVE

The EU institutions have pointed out, in several occasions, an amount of issues concerning the production, sharing, and (re)use of spatial information, which were deemed common to many policy sectors, and transversal to various levels of public authority. Measures addressing the exchange, sharing, access and use of interoperable spatial data and services across different levels of public authority and sectors were therefore reckoned as necessary, and thus the Directive establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) slowly came to be, its purpose to assist policy-making on policies or activities that may have a direct or indirect impact on the environment.

3.3.2.1 HISTORY

Spatial and geographical information is basic for all the activities conditioned by or concerned with specific locations on earth and with their characteristics, be it relating to local, regional, or global scales. The exponential growth of the information that can be collected and processed, and of the means and tools that can be used to process such information, naturally highlighted the necessity of a standardized Spatial Data Infrastructure (SDI) aiming at fostering the sharing of data evaluation in order to eliminate the duplication of the related efforts, at

---

543 Commission of the European Communities, ‘Guidelines for Improving the Synergy between the Public and Private Sectors in the Information Market’ (n 391). See above s. 3.2.1.
546 See INSPIRE Directive, recital no. 4. See also Art. 1.1: “The purpose of this Directive is to lay down general rules aimed at the establishment of the Infrastructure for Spatial Information in the European Community (hereinafter referred to as Inspire), for the purposes of Community environmental policies and policies or activities which may have an impact on the environment”.
making geographic data easily accessible, and at supporting the integration of geographic data from different sources. The state of geographical and spatial data infrastructures in Europe at the beginning of 2000 was indeed deemed in need of a coherent framework, particularly considering the complexity deriving from the multilingual and fragmented European milieu.

The works for the INSPIRE Directive thus started around 2001, and were carried on by the e-ESDI (ESDI standing for European Spatial Data Infrastructure) Expert Group, successively renamed ‘INSPIRE Expert Group’. In December 2001, the Commission issued an ESDI Organisation and E-ESDI Action Plan, which served as basis for the proposal for an E-ESDI framework, which would then become INSPIRE. Individual working groups delved into the most poignant issues foreseen at that time, such as common references for data and metadata, architectures and standards, policy and legal issues, funding, implementation, and the Directive’s foreseen impact and its assessment. As it has indeed been noted, “(t)he development of a spatial data infrastructure (SDI) not only comprises technical aspects but also is supported by economic, social, organizational, and legal measures”.

The following year saw the Commission issuing a memorandum to ensure cooperation between DG Environment, EUROSTAT and the Joint Research Centre (JRC) for the development of the INSPIRE initiative, and E-ESDI expert group releasing a number of position papers on some issues deemed central to INSPIRE’s success. At the beginning of 2003, the Commission launched an open consultation to get feedback on the Directive’s proposal, also shifting its denomination from ESDI to INSPIRE. The proposal was also assessed through an extended Impact Assessment, the same year, making the INSPIRE Directive one of the first EU directives subject to an extended Impact Assessment; notably, the outcome was very positive, highlighting how the proposed INSPIRE Directive would have yielded a return on investment six times the original costs in the environmental sector alone. In 2004, after the processing of the input received from the Expert Group and through the public consultation, the Commission issued its proposal for the INSPIRE Directive. The legislative process followed the ordinary procedure: the initial proposal was amended by both the Parliament and the Council in the two following years, and the contrasts between the legislative bodies were stark.

---

548 Bartha and Kocsis (n 547).
549 Although INSPIRE itself built on the (failed) GI2000 proposal (EC, Towards a European Geographic Information Infrastructure (EGII) (GI 2000); A Discussion Document for Consultation with the European GI Community, 10 May 1995), which aimed at creating the conditions for the development of a European geographic information infrastructure: see van Loenen (n 364) 149.
552 Specifically, on Common Reference Data and Metadata, Architecture and Standards, Legal Aspects and Policy, Funding and Implementation Structures, and Environmental Thematic User Needs.
enough to warrant the adoption of the conciliation procedure, following which the INSPIRE Directive was finally adopted in 2007.\textsuperscript{555}

\subsection*{3.3.2.2 \textsc{Structure}}

The INSPIRE Directive’s purpose is “to lay down general rules aimed at the establishment of the Infrastructure for Spatial Information in the European Community (hereinafter referred to as Inspire), for the purposes of Community environmental policies and policies or activities which may have an impact on the environment”.\textsuperscript{556} It builds upon individual Member States’ infrastructures for spatial information,\textsuperscript{557} in the sense that it does not require the collection of new spatial data.\textsuperscript{558} INSPIRE’s scope and objective clearly touch upon other legislative instruments of the EU, namely the Directive on Access to Environmental Information and the PSI Directive; Article 2 therefore mandates that the INSPIRE Directive’s provisions are to be without prejudice to Directives 2003/4/EC and 2003/98/EC\textsuperscript{559} – the Directive on Access to Environmental Information and the PSI Directive respectively. The Directive, moreover, is not meant to conflict with the existence or ownership of public authorities’ intellectual property rights.\textsuperscript{560} Notably, recital 24 also specifies that the provision of network services within INSPIRE’s scope should comply with the principles relating to the protection of personal data, in accordance with the Data Protection Directive (DPD), thus reminding the interpreter that the INSPIRE Directive’s application has to be applied without prejudice to the DPD, too.

The INSPIRE Directive applies\textsuperscript{561} to spatial data sets, in an electronic format, that relate to an area where a Member State has jurisdiction, and are held by or on behalf of either a public authority (if the data sets have been produced or received by a public authority, or if they are managed or updated by that authority and fall within the scope of its public tasks) or by a third party to whom the network has been made available.\textsuperscript{562} The data sets must moreover relate to one or more of the themes listed in the INSPIRE Directive’s Annex I,\textsuperscript{563} II,\textsuperscript{564} or III\textsuperscript{565}. The INSPIRE Directive applies also to the services built upon the data sets above.

Despite not entailing the collection of data sets \textit{ex novo}, the INSPIRE Directive does require Member States to ensure that metadata is created for the spatial data sets and services falling

---


\textsuperscript{556} INSPIRE Directive, Art. 1(1).

\textsuperscript{557} INSPIRE Directive, Art. 1(2).

\textsuperscript{558} INSPIRE Directive, Art. 4(4).

\textsuperscript{559} See also recital no. 7 of the INSPIRE Directive.

\textsuperscript{560} INSPIRE Directive, Art. 2(2).

\textsuperscript{561} INSPIRE Directive, Art. 4.

\textsuperscript{562} In such cases, when the third party holds IP rights on the spatial data sets, the public authority must obtain the third party’s consent before taking action as prescribed by the INSPIRE Directive.

\textsuperscript{563} Coordinate reference systems, Geographical grid systems, Geographical names, Administrative units, Addresses, Cadastral parcels, Transport networks, Hydrography, Protected sites.

\textsuperscript{564} Elevation, Land cover, Ortho-imagery, Geology.

\textsuperscript{565} Statistical units, Buildings, Soil, Land use, Human health and safety, Utility and governmental services, Environmental monitoring facilities, Production and industrial facilities, Agricultural and aquaculture facilities, Population distribution — demography, Area management/restriction/regulation zones and reporting units, Natural risk zones, Atmospheric conditions, Meteorological geographical features, Oceanographic geographical features, Sea regions, Bio-geographical regions, Habitats and biotopes, Species distributions, Energy resources, Mineral resources.
within its scope, and that those metadata are kept up to date, complete, and of a quality sufficient to describe the relative spatial data sets and services, making it easier (or outright possible) to discover, inventory, and use them. The interoperability and harmonization of the spatial data sets and services is ensured through the adoption of implementing rules, to be enforced by Member States and assessed by the Commission, adopted guaranteeing stakeholder engagement processes. Those implementing rules are meant to address the following themes: a common framework for the unique identification of spatial objects; the relationship between spatial objects; the key attributes and the corresponding multilingual thesauri commonly required for policies which may have an impact on the environment; information on the temporal dimension of the data; updates of the data. The implementing rules are to be designed to ensure consistency and comparability, and have to take into account the relevant standards available.

The INSPIRE Directive’s scope encompasses also the services built upon the geographical data sets considered. In particular, according to Article 11, Member States are bound to create and operate a network of services (discovery, view, download and transformation services, and in general services allowing spatial data services to be “invoked”) and to perform a number of ancillary operations to allow the full development of the network of service. Member States also have to ensure that that public authorities are given the technical possibility to link their spatial data sets and services to the network of service ex Article 11. Derogations from the general regime described are allowed by Article 13, e.g. on the grounds of national security or international relations, confidentiality, intellectual property rights, but they have to be interpreted strictly, as explicitly specified by Article 13’s second paragraph. In general, discovery and view services have to be made available to the public free of charge; however, a derogation is possible for viewing services, in case a fee is deemed necessary for the service’s provision; moreover, the same kind of service can provide data in a form that impedes its commercial re-use. The Commission, finally, is bound by the INSPIRE Directive to establish and operate an Inspire geo-portal at Community level, through which Member States must provide access to their network of services established ex Art. 11.

566 INSPIRE Directive, Art. 5.
567 INSPIRE Directive, Art. 7 ss. See also Article 16.
568 INSPIRE Directive, Art. 7(3).
569 INSPIRE Directive, Art. 7(2).
570 INSPIRE Directive, Art. 7(5).
572 INSPIRE Directive, Art. 8(3) and 8(4).
574 E.g. implementing relevant user requirements, or the combination of search criteria ex Art. 11(2).
576 I.e. the INSPIRE Directive explicitly binds the qualified interpreter to construct its derogatory regime by following what legal methodology historically defines as restrictive interpretation, thus avoiding the possibility of an overly extensive interpretation of the Directive. The provision of direct and binding guidance with regard to the Directive’s interpretation, coupled with the plasticity deriving from the adoption of further implementing rules and with the extensive Impact Assessment and stakeholder consultation, arguably make the INSPIRE Directive an extremely interesting example of modern, sector-specific European legislation.
577 INSPIRE Directive, Art. 13(2).
578 INSPIRE Directive, Art. 15(1).
Member States are obliged to ensure sharing between public bodies as well, enabling them to access, exchange and use spatial data sets and services, for the purposes of public tasks that may have an impact on the environment. Practical restrictions to information sharing are to be eliminated by Member States; however, the use of licenses to allow the use of data sets and services is allowed, and so is the possibility to require a fee, in consideration of the needs of self-sustaining public bodies, but only if compatible with the information-sharing objectives of the Directive, and kept to the minimum necessary to ensure the necessary quality and supply of spatial data sets and services together with a reasonable return on investment. Data and services sharing has to be allowed for other Member States’ public authorities as well, provided there is full reciprocity, albeit Member States are given the possibility to condition their use to requirements foreseen in national legislation. Derogations from the general regime are allowed when data and/or service sharing would compromise the course of justice, public security, national defence or international relations.

Member States are also responsible for setting up the required structures to coordinate the relevant stakeholders at a national level, while the Commission is tasked to do the same at an EU-wide level, with the assistance of the interested organisations, and of the European Environmental Agency in particular.

The INSPIRE Directive has been another major step forward in the development of the European open data policy and legislative milieu. Despite not being often explicitly framed as instrumental for the economic exploitation of PSI, the INSPIRE Directive is certainly an extremely important instrument to this purpose. It highlights how open data is not only about the information per se, but also about the infrastructure through which it is processed. In this sense, and despite the specificity of its subject matter, from a conceptual perspective INSPIRE can be seen as complementary to the PSI Directive – while the latter is mainly concerned with opening up public-sector information, INSPIRE fosters openness at an infrastructural level. However, as highlighted by its content, it would be more fitting to see its provisions as aiming at enhancing transparency and public-sector efficiency, rather than economic growth like the PSI Directive, the other main by-product often associated to the enactment of open data policies and legislation.

3.3.3 THE ACCESS TO ENVIRONMENTAL INFORMATION DIRECTIVE

Both the PSI and the INSPIRE Directives were fundamental instruments in the development of open data and PSI re-use within the EU; albeit having a different scope, they regulate often overlapping subjects, and both their content is – directly or indirectly – directed at opening up data. Another EU legal instrument has had, however, a comparable (even though arguably less stark) impact on the European open data milieu: Directive 2003/4/EC on public access to

---

581 INSPIRE Directive, Art. 17(3).
582 INSPIRE Directive, Art. 17(4) and 17(5).
583 INSPIRE Directive, Art. 17(6).
584 INSPIRE Directive, Art. 17(7).
586 van Loenen and Grothe (n 527).
environmental information\textsuperscript{587} (AEI Directive), one of the main tools through which the Aarhus Convention\textsuperscript{588} has been implemented in EU legislation.

The Directive’s objectives are twofold.\textsuperscript{589} The first is to guarantee the right of access to environmental information held by or for public authorities, and to set out the basic terms, conditions, and practical arrangements for its exercise. The second objective is to ensure that environmental information is progressively made available and disseminated to the public in order to achieve the widest possible systematic availability and dissemination to the public of environmental information, promoting the use of ICT to this end. Accordingly, its Article 3 – titled “Access to environmental information upon request” – requires Member States’ public authorities to make available (as soon as possible)\textsuperscript{590} the environmental information which they held or that is held by a third party on their behalf, to any applicant requesting it, without the applicant having to state a particular interest\textsuperscript{591} in the disclosure. The applicant has the possibility to request the information in a particular format, and the public authority has to comply – unless the data is already publicly available in another form or format, and easily accessible – or if it is “reasonable for the public authority to make it available in another form or format, in which case reasons shall be given for making it available in that form or format”\textsuperscript{592}.

Public authorities, on their hand, must make all reasonable efforts to maintain environmental data in forms or formats that are reproducible and accessible through electronic means, and have also to justify their eventual denial in providing the information requested. Of course, exceptions have been foreseen,\textsuperscript{593} permitting such authorities to decline access in particular cases: however, those exceptions must be interpreted in a restricted way, just like the INSPIRE Directive’s ones. Public authorities can charge for supplying such data, but the amount cannot exceed a “reasonable amount”.\textsuperscript{594} Moreover, they are bound to publish and make available a list of those charges, and additional information on the circumstances in which a charge may be levied or waived. However, if the requests concerns lists of public authorities, or registers or lists of the environmental information held by public authorities or information points, or if the information is examined \textit{in situ}, access must be free of charge.\textsuperscript{595} Access to justice has to be guaranteed to any applicant who considers that his request for information has been ignored, wrongfully refused (whether in full or in part), inadequately answered or otherwise not dealt with; the applicant has therefore to be able to lodge a complaint to another administrative authority, and to have the possibility to review the complaint’s outcome in front of a national court of law.

\textsuperscript{588} i.e. The UNECE (United Nations Economic Commission for Europe) Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters. In particular, the AEI Directive implements (along with the INSPIRE and PSI Directives) the Convention’s first pillar – access to environmental information.
\textsuperscript{589} AEI Directive, Art. 1.
\textsuperscript{590} AEI Directive, Art. 3(2).
\textsuperscript{591} AEI Directive, Art. 3(3). In case an access a request is formulated in an exceedingly general manner, the public authority has to ask the applicant to specify its request, also assisting it in doing so.
\textsuperscript{592} AEI Directive, Art. 3(4).
\textsuperscript{593} AEI Directive, Art. 4 – e.g. when the request is manifestly unreasonable, or when it is formulated in too general a manner, international relations, or again because of public security or national defence, or intellectual property rights.
\textsuperscript{594} AEI Directive, Art. 5.
\textsuperscript{595} Ibid., paragraph 1.
Aside from the information that has to be provided to the user upon request, the AEI Directive also mandates that Member States have to ensure that public authorities organise the environmental information relevant to their functions and held by or for them, and actively and systematically disseminate it to the public, in particular using ICT. Member States have also to guarantee that, over time, the environmental information held by or for their public bodies becomes electronically available. Concretely, such information has to include, as a minimum: the texts of international treaties, conventions, agreements, legislation, policies, plans and programmes that relate to the environment, and the related progress reports; regional and local reports on the state of the environment; data or summaries of data derived from the monitoring of activities affecting (or likely to affect) the environment; authorisations and environmental agreements with a significant impact on the environment; environmental impact studies and risk assessments. Insofar as possible, Member States must ensure that the information above is up to date, accurate, and comparable, when compiled by them or on their behalf.

The AEI Directive is about the right to access to documents, rather than about PSI re-use or open data per se. Its examination is however an interesting peek into some of the precursors of what would later become the idea of open data; amongst others, transparency, access to documents, the belief that some categories of information must be shared by virtue of their social value, and the importance of ICT in ensuring data availability and accessibility.

3.3.4 ON THE LEGISLATIVE SUPPORT TO OPEN DATA

The EU legislation outlined above highlights (some of) the normative backing that the concept of open data has at EU level. Open data is, at its very core, just a set of requirements that aim at making information as accessible, interoperable, and re-usable as possible, by anyone and for any purpose; yet those requirements are expression of an ideal, upheld by a movement, that ended up inspiring concrete policy and legislative change. In turn, the laws examined underline how open data and the many values, ideas, and ideals it aims at advancing require legislative support to be fully accomplished. Open data, and the values underlying it, inspire the adoption of legislation that, in turn, supports their advancement. Indeed, there are several different values underlying open data, and indeed, open data still requires a good deal of legislative backing to overcome the resistance it faces — there is no ‘right to open data’ to speak of. Both those points are expanded upon in the section that follows.

The PSI Directive, the most foundational legislative instrument for the development of the European public-sector information market, fosters open data sharing and re-use at a horizontal level. While it is not meant to be sector-specific, the expansion of its scope over the years (i.e. the 2013 amendments and the recent recast) has added a certain degree of specificity by regulating some sectors (i.e. cultural institutions in 2013, and research bodies and public undertakings operating in the transport and utility sectors with the recast) through

---

596 AEI Directive, Art. 7.
598 Open Knowledge International (n 32).
599 See e.g. Conradie and Choenni, ‘Exploring Process Barriers to Release Public Sector Information in Local Government’ (n 363); Sébastien Martin and others, ‘Risk Analysis to Overcome Barriers to Open Data’ (2013) 11 Electronic Journal of e-Government 348; Barry and Bannister (n 363); Conradie and Choenni, ‘On the Barriers for Local Government Releasing Open Data’ (n 363); Zuiderwijk and Janssen (n 363).
600 See s. 3.4 below, and s. 3.1.2 above.
a special regime of a lower regulatory intensity. Over time, the PSI Directive has indeed shifted from allowing the possibility to re-use PSI to granting the (qualified) right thereof. However, some kinds of data and some kinds of actors have been deemed somewhat ‘sensitive’: while they gradually came to be encompassed within the scope of the PSI Directive, they are still regulated with a lighter touch, and made subject to the previous facultative regime.

Indeed, the scope of PSI legislation and the intensity of the obligation it sets have been growing in parallel with the advance in prominence of the concept of open data on the global policy agenda. The link between PSI and open data has also been made stronger with each set of amendments, to the point that the recast has been renamed to include the words ‘open data’ in its very title. Yet, the origin of PSI legislation in the EU is eminently and inextricably tied with the objective of developing and fostering the (Digital) Single Market, as testified by both the PSI Directive’s history, on one hand, and its structure and legal basis on the other. The concept of open data, on the other hand, supports a broader range of values, rights, and freedoms than just economic growth and development. What the growing convergence between PSI re-use and the concept of open data will lead to remains yet to be seen.

Regardless of whether the market-making rationale underlying PSI legislation and the wide array of other values that the open data movement aims at advancing, the PSI Directive does undeniably foster data sharing and re-use through the removal of both legal and technical barriers. More precisely, the removal or the limitation of those barriers: some norms in the PSI Directive do allow to partly limit, through either technical or legal means, the sharing and re-use of PSI. In this sense, there is a difference between the notion of PSI and the one of open data, in that some limitations that are permissible for the former are not allowed by the latter; not all documents falling under the PSI regime are actual open data, as much as not all open data comes from public-sector sources.

The INSPIRE Directive has been hailed as advancing open data (rectius, PSI re-use) in the EU as well. There are two further insights into the development of open data that have been derived from the brief examination of the INSPIRE Directive above. The first one is the fact that open data initiatives necessitate infrastructural support to reach a level of maturity suitable to the fulfilment of their goals. Indeed, open data seems as much about the contextual and environmental factors enabling data sharing and re-use as it is about the removal of the

---

601 See Janssen and Dumortier (n 388).

602 Article 114 TFEU (formerly 95 TEC), as it concerns the functioning of the Single Market and the free circulation of services.

603 Recital 17 of the PSI Directive (as amended in 2013) states that “In some cases [...] a licence will be issued imposing conditions on the re-use by the licensee dealing with issues such as liability, the proper use of documents, guaranteeing non-alteration and the acknowledgement of source”. In general, as specified by Recital 22, “The [PSI] Directive does not affect the existence or ownership of intellectual property rights of public sector bodies, nor does it limit the exercise of these rights”. Art. 8 mandates that “Public sector bodies may allow re-use without conditions or may impose conditions, where appropriate through a licence”; those conditions, however, may not “unnecessarily restrict possibilities for re-use and shall not be used to restrict competition”.

604 An example of this is given by the (now obsolete) UK PSI Click-Use Licence, which was compliant with EU PSI legislation, but not conformant with the notion of open data (Open Knowledge International, ‘Non Conformant Licenses’ <https://opendefinition.org/licenses/nonconformant/> accessed 28 April 2019).

605 van Loenen and Grothe (n 527).

606 On open data infrastructures, see AMG Zuiderwijk, ‘Open Data Infrastructures: The Design of an Infrastructure to Enhance the Coordination of Open Data Use’ (TU Delft 2015).
information’s legal and technical barriers. The second insight\footnote{607} is that not all data is equally important for the advance of the open data movement and for the development of the European PSI market. Some categories of information are more desirable for re-users, and geographic information is on the top of the list. In this sense, the INSPIRE Directive’s sector-specificity complements well the PSI Directive’s horizontal scope: the two laws are meant to be potentially applicable in parallel.\footnote{608}

Like the INSPIRE Directive, the AEI Directive is concerned with specific categories of information (environmental data) deemed to have particular social value. The AEI Directive can be seen as instrumental to the promotion of open data and information re-use in the EU, in that it fosters the right to access to environmental information, and presents some degree of attention to certain issues that are at the core of the concept of open data, such as technical availability and data quality. Yet, the AEI Directive is about (the right to) access, not about re-use,\footnote{609} and surely not about open data \textit{per se}. Nevertheless, its content delineates (some of) the antecedents of the concept of open data – the value of transparency, and the right to access.

\section*{3.4 Disambiguating open data}

The definition of open data given at the beginning of this chapter is widely accepted,\footnote{610} and yet it is not the only one available in policy and literature – there are many ‘shades of open’,\footnote{611} so to speak. The legislation and policy instruments outlined above are foundational to the development of open data in the EU, but their main subject matters, purposes, and provisions are not really\footnote{612} about open data \textit{per se}, but rather about a disparate array of underlying rights and freedoms; absent a normative definition, which has yet to emerge from law or jurisprudence,\footnote{613} the legal significance of open data is not that clear. This section thus provides

\footnote{607} Which is also confirmed, \textit{inter alia}, by the introduction of the category of ‘high-value dataset’ in the recast of the PSI Directive, and by the identification of particular categories of datasets to be prioritised for release in the European Commission’s Guidelines on recommended standard licences, datasets and charging for the reuse of documents (24.7.2014, OJ C 240/1).


\footnote{609} Yet again, there has always been some overlaps between access and re-use: see K Janssen and J Crompvoets, ‘Access or Re-Use of PSI? A Cookie If You Get It Right!’ \textit{CEUR Workshop Proceedings} (2010).

\footnote{610} The Open Definition, to which this thesis refers when framing open data, and the Open Knowledge Foundation, who formalized the definition, have been explicitly referred to in European Commission, Commission notice — Guidelines on recommended standard licences, datasets and charging for the reuse of documents. 24.7.2014, OJ C 240/1, p. 2.

\footnote{611} Pomerantz and Peek (n 381).

\footnote{612} With the exception of the PSI Directive and of the EC’s communication on open data – European Commission, ‘Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and The Committee of The Regions “Open Data An Engine for Innovation, Growth and Transparent Governance” COM(2011) 882 Final’ (n 386).

\footnote{613} As it has been noted, “a legal definition of what actually constitutes ‘Open Data’ has not yet been provided in either statutes or case law” - Primavera De Filippi and Lionel Maurel, ‘The Paradoxes of Open Data and How
further insight into the ‘plastic’ nature of the concept of open data by highlighting how it is meant to advance a varied assortment of different values, and how it does not, from a teleological perspective, answer to a unitary rationale. The section then clarifies its status vis-à-vis the law, as the idea of balancing open data and data protection that underlies this thesis implies the balancing of rights—and open data is not a right. It concludes by clarifying how open data, in practice, is not a static condition, but rather exists in a spectrum: there are several conceptions and definitions of what constitutes open data, varying amongst each other for the number and stringency of the requirements they set.

3.4.1 THE INSTRUMENTAL NATURE OF OPEN DATA

The sections above dealt with open data—with the notion’s characteristics, genealogy, and development. The examination undertaken, and in particular the analysis of the PSI, INSPIRE, and AEI Directives, highlighted that open data is far from being a unitary concept, but has instead a number of distinct—and yet intertwined—“souls”: access and transparency, PSI re-use and economic exploitation, and public sector efficiency. To those three facets of open data, which have already been recognized by many authors, it would moreover appear prudent to add a fourth one, deriving from a sort of function creep of the declared purposes of open data:

Access and transparency have historically been the first drivers of open data and of its precursors, as highlighted e.g. by the development ‘open government’ movement, or by the chronology of FOIA legislation on both sides of the Atlantic, which predate the instances aiming at fostering PSI re-use for economic reasons, or at increasing public sector efficiency through data-driven governance. The underlying goal is, on one hand, to foster political, administrative, and executive accountability on the authority’s side, and on the other to promote inclusion and participation to democratic processes through the increased availability of the information necessary to partake in decision-making processes. The AEI Directive, as discussed above, exemplifies this kind of antecedent the open data concept, focusing on access and transparency, rather than on the economic growth deriving from data re-use, or on the general efficiency gains deriving from public authorities opening their data.

PSI re-use presents by all evidence another distinct aspiration relatable to the open data movement: the generation of additional value, jobs, and economic growth, through the re-utilization of information openly published by public-sector bodies and other organisations. Navigation systems, interactive library catalogues, weather forecasting services—PSI can be used, if available and accessible, to build products and services from the ground up, potentially giving back to the community of reference the resources spent to gather and process the

to Get Rid of It? Analyzing the Interplay between Open Data and Sui-Generis Rights on Databases’ (2014) 23 International Journal of Law and Information Technology 1, 2. While that is true at an EU level (depending on the outcome of the recast of the PSI Directive), that may very well not be the case in other jurisdictions (e.g. in south-eastern Asian countries) that have not been considered.


Which, incidentally, also applies to the smart city domain, as discussed in the previous chapter.

The value that open data has for entrepreneurship has been held e.g. in Erik Lakomaa and Jan Kallberg, ‘Open Data as a Foundation for Innovation: The Enabling Effect of Free Public Sector Information for Entrepreneurs’ (2013) 1 IEEE Access 558.
information. This is the theme underlying the PSI Directive, as clearly resulting from both its content and its historical evolution, but its traces can also be found in the INSPIRE Directive, which *prima facie* can be seen as rather relating to public sector efficiency and public service provision, but that ended up being an important building block of the EU information market – that very same information market that the PSI Directive aims at enabling.

Open data aims at fostering (public-sector) efficiency as well. The INSPIRE Directive exemplifies such a goal: according to its Article 17, for instance, Member States are bound to adopt measures for the sharing of spatial data between their public authorities to enable them to efficiently carry out their public tasks that may have an impact on the environment, as its goal is “the establishment of the Infrastructure for Spatial Information in the European Community [...] for the purposes of Community environmental policies and policies or activities which may have an impact on the environment”.617 Another example may be the publication of the performance data of publicly funded organisations, both to improve their functioning and to provide the service recipients with transparent information.618 The public sector, in particular, can gain significant efficiency through open data by rationalising data processing, cutting costs and avoiding duplications, and allowing different public-sector bodies, agencies, and departments to re-use the information the others gather.619

The narrative about open data has generally had an extremely positive connotation – transparency, economic growth, and public sector efficiency being clearly desirable goals for everyone.620 However, even though open data and public sector information are usually anonymized before being released, their processing and correlation with other sources may allow to identify sufficiently small subgroups of population to allow the performance of profiling activities, or even to single out specific individuals.621 Open data is a concept, supported by a movement, and rendered concrete by the data infrastructures underlying it, which are open and polyvalent by definition – that is, naturally bound to be utilized by and for a wide number of users and purposes. It would arguably be unrealistic, in view of what has become a widely recognized societal trend, not to consider the potential of open data and of its infrastructures for surveillance and control, whether by design or due to function creep. This appears particularly evident when considering open data in connection with the broader ‘big data’ and ‘big, open, linked data’ (BOLD) buzzwords, whose narratives have been dwindling between an eminently optimistic view622 and more critical and cautious perspectives.623 The

618 Zuiderveen Borgesius, Van Eechoud and Gray (n 41).
620 Although there are, of course negative consequences deriving from open data and PSI re-use: see e.g. Anneke Zuiderwijk and Marijn Janssen, ‘The Negative Effects of Open Government Data - Investigating the Dark Side of Open Data’, *Proceedings of the 15th Annual International Conference on Digital Government Research* (ACM 2014); Peter A Johnson and others, ‘The Cost(s) of Geospatial Open Data’ (2017) 21 Transactions in GIS 434.
622 See e.g. Mayer-Schönberger and Cukier (n 13); Eagle and Greene (n 16).
623 See e.g. Boyd and Crawford (n 159); Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (n 4); Ian Kerr and Jessica Earle, ‘Prediction, Preemption, Presumption: How Big Data
interaction between unprecedented information availability, distributed computing infrastructures, and advanced computational processing and modelling techniques, enables or enhances the potential for surveillance, social sorting, and control, by both public and private actors.

This thesis is concerned with the idea of balancing open data and personal data protection on account of the smart city’s development. It is thus fundamental to remember that open data does not respond to a unitary rationale, but has rather an instrumental nature. Open data is, at its very core, just a set of conditions of a technical and contractual nature that gives substance to the idea that data should be free for everyone to use, modify, and distribute without constraints of any kind. It is important to consider, particularly when open data publishing and re-use compete with conflicting rights and freedoms, and a balancing exercise is needed, what are the actual values, rights, and freedoms that the publication or the re-use instance considered aims at advancing.

3.4.2 NOT A RIGHT (ON ITS OWN)

One of the assumptions underlying the research project from which this thesis is the necessity, or perhaps the mere desirability, of a balancing framework mediating between the right to personal data protection and the interests underlying open data. Intuitively, and broadly speaking, this framing makes sense: there is a fundamental tension between openness and protection, and finding an optimal trade-off between the two seems like the most rational course of action. The concept of ‘balancing’ is discussed in chapter 5, after laying down this thesis’ theory of data protection in chapter 4. A preliminary step is to clarify the legal significance of the notion of open data: the idea of ‘balancing’ and the growing prominence of the concept of open data in the political agenda may engender the idea that open data and personal data protection should be considered “on an equal footing”, while they should not.

As opposed to data protection, open data is not a right, let alone a fundamental right. Rather, the most fitting way to see open data is as a set of requirements constituting the by-product of a social movement and of an ideal – one of transparency and efficiency and serendipitous discovery of more value in government data. Such movement, and the ideas on which it is based, have in turn been leading to the adoption of policies at all level of government, at international level, and within individual organisations in the public and even in the private sector. Sometimes, the ideas underlying open data are also transposed in some provisions of certain laws, such as the PSI and INSPIRE Directives discussed above, which mandate some obligations and grant some rights aiming at enhancing public-sector data sharing and interoperability. That does not however mean that open data, as a whole, can be seen a right in itself – just that the ideals on which it is based, some of which are fundamental rights, have been made more concrete and institutionalised through normative means.

An example can be given through the ‘right to re-use’ granted by the PSI Directive after its 2013 amendments. Initially, the 2003 PSI regime, which granted a faculty to re-use documents, according to which Member States had to ensure that, “where the re-use of


624 See s. 3.2.1 and 3.3 above.

625 E.g. the right to access to documents, freedom of business, freedom of information, and the right to good administration.

626 See above s. 3.3.1.
documents held by public sector bodies is allowed, these documents shall be re-usable for commercial or non-commercial purposes” in accordance with the conditions set out by the remainder of the PSI Directive. The 2013 amendments to the PSI Directive modified such a faculty into a right – the general principle now states that “Member States shall ensure that documents to which this Directive applies in accordance with Article 1 shall be re-usable for commercial or non-commercial purposes”, again in accordance with the conditions set out by the remainder of the PSI Directive.\textsuperscript{627} In a nutshell, if the data is there, no exemption applies, and the public-sector body subject to the PSI Directive\textsuperscript{628} does not have to create the dataset requested \textit{ex novo},\textsuperscript{629} potential re-users have a positive legal right\textsuperscript{630} to obtain the data in consideration, within the limitations set by the rest of the PSI Directive. \emph{That} – re-use – is a right, delimited and defined by its boundaries. Open data, in this sense, is rather the ideal to which the right to re-use can be seen as instrumental to.

In turn, open data bundles a set of different underlying values (e.g. transparency, accountability, growth, efficiency) which are quite divergent amongst each other, so much that they might even conflict. The issue is not one of specificity: the concept of open data, if not diluted, can be fairly specific. Rather, it is about open data’s partial overlaps with other concepts (and well-established rights), such as access, good administration, and freedom of information, and (thus) about its lack of autonomy. Open data is a data publication and re-use policy\textsuperscript{631} suitable to advancing those rights and freedoms, but it is not the only way – it is not strictly \textit{necessary}, in other words – and, when the rights and freedoms underlying it happen to conflict with countervailing rights and freedoms, is seldom a proportionate answer. This will be discussed further in chapter 5, in respect to the idea of balancing open data and data protection at a fundamental rights level.

Open (government) data rather belongs to the policy realm, to the executive domain more than to the legislative one. Its function is instrumental to the pursuit of rights and values, but that does not make it one of them \textit{per se}. A similar point has been made by Yu and Robinson,\textsuperscript{632} who point out how there may be a gap between the data that actually makes governments (or organisation) more ‘open’ in the sense of transparent and accountable, and the publication of “politically neutral” information that is easy to access and re-use, even if it has no direct link with public accountability. Equating the concept of open data with its objectives, and with the rights and values underlying them, looks akin to conflating the end with the means, or to mistaking the trees for the proverbial forest. This will remain a basic postulate throughout the

\textsuperscript{627} 2013 PSI Directive, Chapters III and IV.  
\textsuperscript{628} The general principle of the PSI Directive, as amended in 2013 and as maintained by the 2019 recast, does not apply to all kinds of public-sector bodies.  
\textsuperscript{629} See Art. 4 of the PSI Directive.  
\textsuperscript{630} Positive rights oblige certain subjects – the government, private individuals, legal persons – to provide benefits to the rightsholder; they differ from negative rights, which forbid the negation of certain actions or properties to the rightsholder. Often, provisions that grant a positive right also grant a complementary negative right – in the case of the right to re-use granted by the PSI Directive, the positive right is the right to be provided with the PSI requested to a public-sector body; the negative right is not to have a legitimate request prevented or denied. Legal rights are granted by statutory law, as opposed to contractual rights, which are granted through contractual means and backed up by the legal rights to make agreements and to have them enforced. See Fred E Foldvary, ‘Positive Rights’ in Deen K Chatterjee (ed), \textit{Encyclopedia of Global Justice} (Springer Netherlands 2011).  
\textsuperscript{631} \textit{Rectius}, a set of requirements and licensing conditions that substantiates a data publication and re-use policy that is, sometimes, backed up by legislative means.  
\textsuperscript{632} Yu and Robinson (n 384).
thesis: the issue is not the necessity of balancing open data and data protection, but rather balancing the rights and freedoms underlying open data with the right to data protection. Whether that is possible or desirable is discussed in the following chapters.

3.4.3 DEGREES OF OPEN (DATA)

Open data is constituted, at its very core, by a set of high-level requirements aiming at making certain information as accessible, interoperable, and reusable as possible. One of the characteristics that are most noticeable about the concept of open data is how and how much the number and intensity of those requirements changes according to the definition, policy, or document considered. Indeed, in many jurisdictions, “a legal definition of what actually constitutes ‘Open Data’ has not yet been provided in either statutes or case law”.\(^{633}\) That, as it has been pointed out, is an issue that pertains more generally to the meaning of ‘open’, be it in open data, open source, or open government, to name a few: “(u)ntil quite recently, the word “open” had a fairly constant meaning. The over-use of the word “open” has led to its meaning becoming increasingly ambiguous. This presents a critical problem for this important word, as ambiguity leads to misinterpretation”.\(^{634}\)

Such ambiguity is not a mere semantic issue. Openness is a concept that is increasingly being translated into policies and legislation at a local, national, and international level. Open data’s actual capacity to foster innovation, growth, transparency, accountability, and efficiency, depends on how the concept is framed and interpreted. Furthermore, open data is also a movement, and comes with its (lato sensu) political agenda. Framing a government branch or a data sharing policy as ‘open’ carries a narrative value independently of how (and perhaps even whether) the underlying requirements are actually implemented. The need for a rigorous definition (and conception) of open data has already been underlined\(^{635}\) as serving the function of quality control, in that it avoids “open-washing” and the dilution of the meaning of ‘open’. It also boosts the compatibility of different open data sets from both a legal and a technical perspective, rendering the conditions required e.g. for licensing or technical interoperability between datasets homogeneous. Finally, an agreed-upon definition fosters simplicity – it is easier to talk about open data without having to clarify the boundaries of the concept beforehand, defining less permissive data sharing policies with different terms.

More specifically to this thesis, another observation emerged during the moments of engagement with the relevant stakeholders. Many of the mechanisms that were highlighted as potential solutions to “balancing” open data and data protection (e.g. access controls, use restrictions, licensing conditions, shifting to interactive systems altogether) were often deemed as providing an adequate trade-off between the rights, freedoms, and interests involved; yet many stakeholders also reacted (or outright objected) with something to the tone of “but that would not be open data anymore!”\(^\)\(^{636}\). Being able to label a particular program or policy or single data release as ‘open data’ seems to have some sort of narrative value on its own. Yet, there is an undeniable degree of disparity between the number and rigorousness of the requirements set by the various open data definitions available in literature.\(^{636}\) Many

---

\(^{633}\) De Filippi and Maurel (n 613) 2. I understand that not to be the case everywhere.

\(^{634}\) Pomerantz and Peek (n 381).


\(^{636}\) See Annex I at the end of this thesis.
definitions insist on the technical side as much as on the lack of intellectual property limitations, requiring full machine-readability, open standards, and open formats before defining accessible information as ‘open data’. Some others acknowledge how openness is a spectrum, rather than a static condition, so that data policies and initiatives can be seen as being ‘open’ from and up to a certain degree of maturity. In practice, the open data policies in which those definitions are translated do not always result in the publication of fully open data – a number of practical constraints often result in information made available under restrictions that still somehow limit its accessibility, usability (by anyone and for any purpose), or interoperability.

In any case, this thesis moves from the assumption that “open means open”, and that there are other words that can be used to define data sharing endeavours that are more permissive than just keeping the information in-house, but that do not present the same level of attention to factors other than accessibility, such as e.g. interoperability, reusability, and the other underlying freedoms open data is based on. Accordingly, the dissertation refers to open data in a strict sense, as defined in its initial sections, and uses a different expression – PSI, data sharing, and the likes – where more precise and appropriate. Some overlaps are however inevitable: on one hand, the terminology often overlays; on the other, sometimes what is dubbed ‘open data’ is not entirely open, but rather only ajar or agape.

3.5 CONCLUSION

This chapter delved into the concept of open data, briefly accounting for its recent developments and constitutive elements. With regard to the latter, it is important to note how there are many degrees of openness, and not all of them can (or should) be defined as actual open data: let us not dilute the meaning of a term that engendered a whole movement. In a way, calling any permissive data release ‘open data’ would be close to equating open source software to ‘freeware’ – there is much more to it than mere price or accessibility.

The sections above outlined the requirements that are generally understood to integrate the definition of open data, the absence of which places the information considered on another shade of the openness spectrum, without however rendering it necessarily ‘closed’, inaccessible or unusable. They also discussed (some) of the EU laws and initiatives that provided a regulatory backbone for open data publication and re-use, the most significant of which is the PSI Directive. Yet, PSI is not really tantamount to open data. On the one hand, PSI law has an eminently economic underpinning, while open data has a broader set of goals: aside from fostering innovation and growth, open data is also concerned with increasing transparency, accountability, and (public-sector) efficiency. On the other hand, PSI legislation does not impede setting conditions, such as access controls and use restrictions, that are incompatible with open data requirements.

---

637 E.g. Open Knowledge International (n 32).
639 At an organizational level, developing, implementing and maintaining open data programs requires a certain amount of time and resources.
640 E.g. tabular data shared in PDF format, or information that is accessible, re-usable, and interoperable, but covered by a license that is not open.
It is important to acknowledge how different ‘souls’ live within the notion of open data: one that aims at bolstering transparency and accountability, another that sees open data as a tool for innovation and growth, and a last one framing it as a tool to foster governmental efficiency.\textsuperscript{641} That is meaningful since open data is just a data sharing policy, not a right in itself: it can be \textit{used} to uphold individual rights and freedom, but does not constitute a right on its own. The rights a particular open data release is meant to uphold, in turn, can be in conflict with other countervailing rights, such as privacy and data protection, and is thus important to understand which rights an open data sharing instance is meant (or rather, more likely) to support. That is to say that it is not open data \textit{per se} that needs to be balanced with individuals’ right to personal data protection, but rather the rights and freedoms (e.g. freedom of information, freedom of business, the right to access, and so forth) underlying the specific open data release or re-use instance in consideration.

The concept of open data – be it the requirements in which the overarching idea substantiates, the laws and policies supporting it, or (to some extent) the movement pushing for it – is inherently instrumental in nature. That is, as Chapter 5 lays out, a major consideration to be kept in mind when investigating the idea of balancing open data and data protection.

\textsuperscript{641} See Zuiderveen Borgesius, Van Eechoud and Gray (n 41); Scassa (n 614).
4. A THEORY OF DATA PROTECTION

The second chapter of this dissertation examined the concept of smart city, concluding that it is, for the better part, a social construct – a plastic notion fit to accommodate a broad set of narratives. It is, however, a socio-technical construct,642 underpinned by a wide array of information and communication technologies that are inherently threatening to individuals’ privacy and data protection rights. Rectius, those technologies can be inauspicious to privacy and data protection – lots depends on their concrete design and implementation.

The thesis then dealt with open data, on the assumption that it has a role to play in the development of the cities of tomorrow;643 it underlined what open data is – a movement based on ideas that are increasingly being translated into law and policies, but not a right on its own. The dissertation followed up by placing data protection within the broader set of barriers faced by the open data movement; it also highlighted the impact that open data can potentially have on individuals’ rights to privacy and data protection by examining how the issue is dealt with by policy and legislation.

The other side of the coin – the impact that data protection law can have on open data publication and re-use initiatives is discussed in the remainder of the thesis. Before delving into how (and whether) data protection curtails open data, it seems necessary to discuss what data protection actually is – its development, rationale, and essential traits. The stakeholders engaged and the literature surveyed do indeed often conflate the right to privacy with the one to data protection, lamenting that EU data protection law limits the possibility to share as open data information that is not private under the conventional meaning of the term. That may very well be so, but it is also immaterial: privacy, or the right to the respect for private life, is not tantamount to the right to the protection of personal data.

The European right to personal data protection is still a young right. While it is still often equated to privacy, this chapter argues that it rather came to constitute an entirely different regulatory and societal response to the growing importance and pervasiveness of computing in the modern society, to the same impulse that gave rise to the instrumentation of the built environment, and the datafication of society.644 Data protection is historically entwined with (informational) privacy. Decades ago, the birth of mainframe computing and databanks spurred a broad debate about the perils of their misuse,645 and about the drawbacks of technological innovation as applied to data processing in general. Now, in the smart city, computer code and the built environment merge646 in a multi-layered social and technological construct, embedded within a much broader system of systems. Ubiquitous computing, the cloud, artificial intelligence, and big data analytics, have fundamentally altered the scale and scope of modern data processing activities. The need for restraints, responsible innovation,

642 Sadowski and Bendor (n 226).
644 See chapter 2.
646 See Lessig, Code and Other Laws of Cyberspace (n 67); Lessig, Code Version 2.0 (n 67); Kitchin and Dodge (n 10); Benjamin H Bratton, The Stack: On Software and Sovereignty (MIT Press 2016).
and fairness in information processing remains substantially the same. Yet, such change in scale and scope questions the adequacy of the regulatory frameworks that were devised to curb power and information asymmetry between the mainframe computing era and the start of the personal computing one. The smart city – the pervasiveness and ubiquity of data collection and processing technologies embedded within the built environment – challenges the traditional construction of data protection.

At the same time, the global strive towards openness sketched in the preceding chapter is on the rise, facilitated by ICT developments. Open data has been framed as a major building block of the cities of tomorrow, but it is also intrinsically at odds with the disclosure limitations sanctioned by data protection regulations. This dissertation’s purpose is to examine, from a regulatory perspective, the tension between data protection and openness in the context of the instrumentation and datafication of the built environment. This chapter thus deals with the right to data protection (and, incidentally, with the one to privacy) by outlining this thesis’ theory of data protection, conjecturing its substance and essence.

The chapter is structured as follows: the following section explains the materialisation of the right to data protection in the EU, starting from its beginnings under the aegis of the right to privacy’s and ending with the current status quo: its (quasi)constitutionalisation by the Charter and the Lisbon Treaty, and the recent reform of EU data protection law. The second section distinguishes the rights to privacy and data protection by reference to their horizontal and vertical effect, their scope, the kind of (sub)rights they grant, and their societal function. The third section outlines this thesis’ framing of the substance of the right to personal data protection, and the fourth section discusses its essence. The fifth section considers the implications of this thesis’ theory of data protection for open data publishing and re-use.

4.1 ON THE MATERIALISATION OF THE RIGHT TO PERSONAL DATA PROTECTION

The emergence of the European right to the protection of personal data has been a bumpy road. Its process of (quasi)constitutionalisation – arguably a sui generis one – and its history and development offer an interesting window on the EU’s strategy on the regulation of (personal) information, both from a fundamental rights and from an internal market perspective. Data protection is highly instrumental to the protection of all the rights and interests that can be dented by unfair information processing, and provides for one of the main regulatory frameworks controlling data flows both within the European (Digital) Single Market and, to a certain extent, outside of it. It serves as an external, normative boundary to both the pervasiveness of the instrumentation of the built environment, and to open data initiatives in general.

Balancing the values underlying open data with the right to personal data protection to foster the development of the smart city environment requires the understanding, in addition to the law’s application and mechanics, of the ratio underlying such regulatory barrier to information

---

647 See Ceruzzi, Computing: A Concise History (n 99); Ceruzzi, A History of Modern Computing (n 99).

648 See e.g. Walravens, Breuer and Ballon (n 14); Hendrik Hielkema and Patrizia Hongisto, ‘Developing the Helsinki Smart City: The Role of Competitions for Open Data Applications’ (2013) 4 Journal of the Knowledge Economy 190; Martin Bartenberger and Verena Grubmüller, ‘The Enabling Effects of Open Government Data on Collaborative Governance in Smart City Contexts’ [2014] Available at SSRN 2474974.
processing. The coming subsections therefore deal with the materialisation of data protection within the EU’s fundamental rights framework, and with its relationship with the right to privacy, from which the former stems. To begin, I will briefly account for the birth of the right to privacy, and for the development of the so-called “computers and privacy” issue, foundational to the birth of the concept of personal data protection. The following sections describe the emergence of the right to data protection in Europe both from a national and from a supra-national perspective, tracing it to the US “computers and privacy” matter. The aim of this section is to explore the birth and rationale of the European right to personal data protection, introducing its (admittedly foggy)\(^{649}\) distinction from the right to privacy, in order to frame its position and hierarchy in respect to the concept of open data, and its role as a safeguard from the unrestrained instrumentation and datafication of the built environment – the so-called “smart city”.

4.1.1 THE RIGHT TO PRIVACY

The right to privacy, despite having arguably more far-reaching antecedents, is traditionally led back to Warren and Brandeis’ homonymous article,\(^ {650}\) itself a reaction to the perceived intrusiveness of that time’s press \(^ {651}\) and to the development of the photo camera technology.\(^ {652}\) The accounts of the reasons that spurred the two Boston lawyers to write “The Right to Privacy” have been widely debated, but it seems overall that the “ample coverage of the Warrens from 1883 to 1890, regaling readers with breathless accounts of their weddings, social gatherings, and funerals, […] very plausibly could explain Warren’s evident desire in "The Right to Privacy" to rein in the press through new tort protection for personal privacy”\(^ {653}\).

Privacy, from the start, has been meant to curb information availability and disclosure, and has been – at least partly – spurred by technological development. Despite the natural dichotomy between public and private, inner and outer, seclusion and publicity, before 1890 “there was no compelling call for a sector in the law that recognized the “right of privacy” as one of these basic rights and the need to protect it legally”\(^ {654}\). Granted, several normative provisions – in the US and elsewhere – already safeguarded individuals’ “right to be left alone”, as the right to privacy was initially conceptualised: professional secrecy, for instance, or laws on trespassing, or on the confidentiality of communications,\(^ {655}\) largely predated the emergence of the right to privacy. However, the 1890 Warren and Brandeis’ article\(^ {656}\) has been the main steppingstone and starting point for the birth of privacy as a

---

\(^{649}\) See e.g. M Tzanou, ‘Data Protection as a Fundamental Right next to Privacy? “Reconstructing” a Not so New Right’ (2013) 3 International Data Privacy Law 88, 88; Gloria González Fuster, ‘Fighting for Your Right to What Exactly - the Convoluted Case Law of the EU Court of Justice on Privacy and/or Personal Data Protection’ (2014) 2 Birkbeck L. Rev. 263.

\(^{650}\) Warren and Brandeis (n 258).


\(^{653}\) Gajda (n 651).


\(^{656}\) Warren and Brandeis (n 258).
standalone right. It would be however substantially reductive and historically imprecise to reduce the genesis of the right to privacy to the brainchild of two young Boston lawyers. The values underlying the right to privacy, from the second half of the 19th century, increasingly came to be threatened by social and technological change, and led to a public push for privacy regulation.\textsuperscript{657} Innovation – the telegraph and the photo camera and the “yellow-page” press – gave rise to a plethora of issues who called for a normative response. Such a response was the birth of privacy.

The right to privacy – and the torts to which its breach lead – had a slow emergence at first: Warren and Brandeis’ article had initially little impact in US judicial practice, and for the following few decades the very existence of the right to privacy was debated by US courts. In the thirties, due to the publication of the first Restatement of Torts,\textsuperscript{658} the right to privacy came to be increasingly recognised by the courts, and the decisions rejecting its existence started being overruled. By 1960, it was overwhelming recognized by the American courts,\textsuperscript{659} at least in one of its forms. As it had been put by authoritative doctrine,\textsuperscript{660} by 1960 US privacy came to comprise four different torts deriving from the invasion upon one’s seclusion, in a taxonomy that shaped American understanding of privacy for the following decades.\textsuperscript{661}

1. Intrusion upon seclusion, solitude, or private affairs.
2. Public disclosure of embarrassing, private facts.
3. Publicity which places the plaintiff in a false light in the public eye.
4. Appropriation of one’s name or likeness.

In the meantime, WWII’s aftermath\textsuperscript{662} saw the recognition of the right to privacy at an international level.\textsuperscript{663} Of paramount importance, in this respect, is the foundation of the Council of Europe (CoE) in 1949 and to the subsequent drafting of the European Convention on Human Rights (ECHR) in 1950.

The right to privacy is incorporated in Art. 8 ECHR,\textsuperscript{664} which states that “Everyone has the right to respect for his private and family life, his home and his correspondence”, and mandates that “There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention

\begin{thebibliography}{99}
\bibitem{657} Shank (n 654).
\bibitem{658} “Within 50 years, the Restatement of Torts, reflecting the law in most U.S. jurisdictions, considered invasion of privacy to be a cause of action”: Patrick E Cole, ‘New Challenges to the U.S. Multinational Corporation in the European Economic Community: Data Protection Laws ’ 893 895.
\bibitem{659} Prosser (n 260).
\bibitem{660} Prosser (n 260).
\bibitem{662} “The political philosophy and human rights agenda of the Allied Powers after the Second World War shaped both the selection of rights protected and the machinery for their protection” – Bernadette Rainey, Elizabeth Wicks and Clare Ovey, \textit{The European Convention on Human Rights} (Bernardette Rainey, Elizabeth Wicks and Claire Ovey eds, 7th edn, Oxford University Press 2017) 3 ss.
\bibitem{664} Rainey, Wicks and Ovey (n 662) 400 ss.
\end{thebibliography}
of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others”. While the ECHR does not contain any article specifically dedicated to data protection, over time Art. 8 has gradually been interpreted by the European Court of Human Rights (ECtHR) as covering personal data processing as well. Indeed, right since the promulgation of the ECHR, the right to respect for private and family life has been much broader than mere privacy as the right to be left alone, to live as one wishes, far from the public’s eye. Through the creative, evolutionary jurisprudence of the ECtHR, Art. 8 has been used to protect an extremely broad set of goods and values, such as freedom from interference with one’s physical or psychological integrity, searches of property and electronic surveillance, the protection of and the freedom to develop one’s identity, immigration, health, medical procedures, and reproductive rights, and the protection of one’s living environment and home. This brief review is by no means exhaustive, and the wide-ranging assortment of rights encompassed by Art. 8 ECHR’s umbrella, as it has been noted, “may not yet be fully explored”.

Privacy has thus been, from its very start, extremely multifaceted, and highly connected to technological evolution and to its societal impact. In the first half century following its

---

665 ECHR, Art. 8 – Right to respect for private and family life.
666 Compulsory medical interventions, for instance, have been found to be a compression of Art. 8 ECHR (ECtHR, YF v. Turkey, 39 E.H.R.R. 34 (2004)), and likewise e.g. inappropriate searches of prison visitors (ECtHR, Wainwright v. United Kingdom, 44 E.H.R.R. 40 (2007)) and police “stop and frisk” powers (ECtHR, Gillan and Quinton v. United Kingdom, 2010 E.C.H.R. 28, 50 E.H.R.R. 1105 (2010)), as they constitute a compression of physical integrity (and thus of the right to private life). The same goes for psychological integrity (ECtHR, Y v Slovenia App. No(s). 41107/10 (2015)) and the recording and dissemination of images (ECtHR, Von Hannover v. Germany, 40 E.H.R.R. 1 (2005)).
667 E.g. ECtHR, Niemietz v. Germany (n 300).
669 E.g. in the context of the search for parental identity (ECtHR, Odievre v. France, 38 E.H.R.R. 43, 2003 F.C.R.1 621 (2004)), or to safeguard one’s cultural or ethnic identity (Buckley v. United Kingdom, 23 E.H.R.R. 101 (1996); Ciubotaru v. Moldova, App. no. 27138/04, (2010)). A person’s name falls under the protection of Art. 8 ECHR too (see e.g. ECtHR, Stjerna v. Finland, 24 EHR 195 (1997)), and so does his or her sexuality (ECtHR, Dudgeon v United Kingdom, 4 E.H.R.R. 149 (1981); Laskey, Jaggard and Brown v. United Kingdom, 24 E.H.R.R. 39 (1997)).
671 Art. 8 ECHR has been invoked e.g. in cases revolving around consent for medical procedures (e.g. ECtHR, Glass v. United Kingdom, 39 E.H.R.R. 15, 2004 F.L.R.1 1019 (2004); VC v. Slovakia, 2011 E.C.H.R. 1888 (2011)), reproductive rights (Tysiak v. Poland, 45 E.H.R.R. 42 (2007); A, B and C v. Ireland, 2010 E.C.H.R. 2032, 53 E.H.R.R. 13 (2010)), and euthanasia (Pretty v. United Kingdom (n 300)).
672 As for the former, the ECtHR has examined environmental considerations under Art. 8 e.g. in ECtHR, Kytatlos v. Greece, 40 E.H.R.R. 16 (2005); Guerra and Others v. Italy, 26 Eur. Ct. H.R. 357, 26 ECHR 3577 (1998). As for the protection of the home, see e.g. ECTHR, Gillow v. United Kingdom, 11 E.H.R.R. 335 (1986).
673 Rainey, Wicks and Ovey (n 662) 454.
674 And still is: see e.g. Solove (n 252); Koops and others (n 263); Rachel L Finn, David Wright and Michael Friedewald, ‘Seven Types of Privacy’ in Serge Gutwirth and others (eds), European Data Protection: Coming of Age (Springer 2013).
conceptualization as a standalone right, it factually developed into a system of diverse torts related by their impact on an individual’s intimate sphere. From the second half of the 20th century, privacy began intersecting with a novel technological development – the computer – forming the basis for the birth of the concept of data protection.

4.1.2 COMPUTERS AND PRIVACY: FROM THE US TO THE EU

From the 1960s onwards, another side of privacy rose to public attention on both sides of the Atlantic: the issue of computers and privacy, arguably the very inception of the genesis of data protection as an individual right within the EU’s (quasi)constitutional framework. Balancing countervailing rights and interests in a highly technological environment – data protection and open data in the smart city – requires identifying data protection’s underlying rationale, and its development over the past decades, which is historically entwined with the right to privacy.

4.1.2.1 COMPUTERS AND PRIVACY IN THE US

From the end of the eighteenth century, several social processes involving the political West – most significantly the UK, the US and continental Europe – fostered the need for centralized information collection. Amongst other factors, one can identify the advances in communication and transportation technologies; the growth of the active electoral pool; the rampant urbanisation deriving from the growing industrialisation and consequent migrations from the countryside to the city, and thus an increasingly condensed workforce. 675 “(B)ureaucracy emerged as a response to nineteenth-century industrialization”. 676

During the second world war, military necessities spurred – on both sides of the conflict – a wave of research and development initiatives aiming at creating technologies that would allow the nations involved to get the upper hand in the escalating confrontation. From the code-breaking efforts of Bletchley Park,677 to the development of semi-autonomous anti-aircraft and anti-missile systems to counter German ballistic progress, 678 war spurred a wave of innovations. After the end of the conflict, part of the investment and the technologies that were designed or employed for military purposes were adapted to peacetime needs, resulting, at the beginning of the ’50s, in the start of the mainframe era. “The computer, originally designed to track missiles and break secret codes, came at exactly the right time for government bureaucracy”.679

The two decades between 1950 and 1970 saw the rise of general-purpose mainframe computers. By 1960, roughly 3500 mainframe computers operated within the US: large arrays of circuitry mounted on large metal frames,680 kept in climate-controlled rooms, and operated by a highly specialised techno-clergy, able to make sense of the arcane assemblage of magnetic tapes and punch-cards through which mainframes functioned.681 Mainframes were a tool of the few – complex to operate and maintain, and extremely expensive. Used to manage payrolls

---

676 Weller (n 675).
680 Hence the name: mainframe – Ceruzzi, Computing: A Concise History (n 99) 54.
681 Rid (n 678) 67–68.
and to process census data, to crunch numbers for taxation and to store records about the population, mainframe computers brought along an enhancement of the power and information asymmetry running between the subjects to which the information processed relates and the actors that control the computing.

Concerns about the dark side of data processing were not new – the census itself has been an historically problematic instrument, and so have been e.g. the postal system, the telegraph, the telephone and its wiretapping, the photo camera, and the growth of government record systems in general. The magnified information storage and processing capacity deriving from the rise of mainframe computing caused however newfound concern within a number of experts. As reported by González Fuster, Bernard S. Benson – founder of a Californian niche technology company – claimed already in 1961 that the growing amount of personal data stored in computers could one day be fed into a single system, with clear implications for individuals’ privacy.

A few authors, in the following decade, drew the link between computing power, data processing, and individuals’ rights and freedoms, most prominently privacy. In 1964, Packard published “The Naked Society”, a timely early warning about the growing dangers posed by the rising information society. 1967 saw the publication of one of the most important texts on privacy (and data protection) of all times, possibly right after Warren and Brandeis’ seminal paper: Alan Westin’s “Privacy and Freedom”. Westin’s book is concerned with the threats to privacy deriving from computerisation, and puts forth a new definition of privacy as the right to determine autonomously when, how, and to what extent one’s information can be communicated to others: the roots of privacy as informational self-determination.

Those years saw also the involvement of US governmental and legislative bodies in the development and modernisation of the concept of privacy. In 1965 and 1966, the US Congress’ House Special Subcommittee on Invasion of Privacy of the Committee on Government Operations began discussing and holding hearings on the linkage between computers and privacy in the context of a discussion revolving around the creation of a federal, centralised databank. In 1971, reportedly in part due to Arthur Miller’s book “The Assault on Privacy”, the Subcommittee on Constitutional Rights of the Senate began holding hearings on the matter as well. In 1972, the Secretary of the US Department of Health, Education and Welfare established an Advisory Committee on Automated Personal Data Systems, which

682 See e.g. Desai (n 655); Solove (n 652).
684 Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 28. The seminal warnings of US early computer experts are also mentioned in David P Farnsworth, ‘Data Privacy or Data Protection and Transborder or Transnational Data Flow, an American’s View of European Legislation’ (1983) 11 International Business Lawyer 114.
686 Warren and Brandeis (n 258).
687 Westin (n 242).
688 Known as the ‘Gallagher committee’ after its chair, Cornelius E. Gallagher.
689 Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 30.
690 Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 33.
691 Miller, The Assault on Privacy: Computers, Data Banks, and Dossiers (n 242).
produced, in 1973, a report titled “Records, Computers, and the Rights of Citizens”. The report addressed privacy concerns in personal data processing by suggesting a number of fair information processing principles, and advocated for the creation of a code of ‘fair information practice’ regulating governmental automated personal data processing. Fair information processing principles and practices (oftentimes referred to as FIPPs) became since then a staple component of data protection legislation worldwide. In 1974, after the establishment of the Domestic Council Committee on the Right of Privacy, the US enacted its first federal privacy law – the US Privacy Act – itself heavily reliant on the FIPPs doctrine.

4.1.2.2 Computers and Privacy in Europe

The seminal writings and warnings of US computer experts and scholars – Westin above all others – echoed heavily across the pond. At the same time, the European population – as happened in the US – was gradually growing wary of the fact that computers magnify power and information asymmetries, thus potentially posing a serious threat to privacy. The European regulatory reaction, which started between the ‘60s and the ‘70s, formed the basis for the genesis of the right to personal data protection. From the end of the ‘60s, echoing the American initiatives, European scholars began exploring possible regulatory solutions to the potential invasions upon one’s privacy, in particular if deriving from modern computing technologies.

The International Commission of Jurists (ICJ), an international human rights NGO, for instance, was a very active actor in this respect. In 1967, it organised the Nordic Conference of Jurists on the Right to Respect for Privacy, from which a report on the modern concept of privacy was produced. In 1970, following the 1968 UN International Conference of Human Rights in Tehran and an UNESCO expert meeting on human rights and new technologies, UNESCO entrusted the ICJ with the publication of another report. In the same year, the British section of ICJ (JUSTICE) thus published a report titled “Privacy and the Law”, where it also explored the issue of computers and privacy. In 1972, the ICJ published its final study, which placed lots of attention on the threats posed by computerised data processing. The study was largely influenced by Westin’s book, and by its notion of privacy as control upon personal information.

694 "In the United States, which is ahead of Europe in its computer development, social scientists have analysed this social problem": Frederik Willem Hondius, Emerging Data Protection in Europe (North-Holland Publishing Company 1975) 3. “A factor of considerable influence was the development of data protection on the American scene”: Hondius, Emerging Data Protection in Europe 6. See also Adam Warren and James Dearnley, ‘Data Protection Legislation in the United Kingdom’ (2005) 8 Information, Communication & Society 238.
695 See Hondius, Emerging Data Protection in Europe (n 694); Mayer-Schönberger (n 679).
696 Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 39.
697 “The "data revolution," by strange historical coincidence, began in Tehran in 1968, ten years before the Iranian revolution. The setting was the International Conference on Human Rights called by the United Nations on the 20th anniversary of the Universal Declaration of Human Rights”: Frits W Hondius, ‘Data Law in Europe’ (1980) 16 Stanford Journal of International Law 87, 90. See also Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 40.
At the same time, national actors were exploring the topic as well. After the publication of JUSTICE’s report, which endorsed the formalisation of a general right to privacy in national law, a few related legislative proposals were attempted, even if none of them ever saw the light. In 1970, the UK government created a Committee on Privacy (the Younger Committee). In 1972, the Younger Committee published its report, which formulated a number of principles relating to the processing of personal data through computerised means. In 1975, following the Younger report, the UK government published two white papers on computers and privacy, and set up a Data Protection Committee, chaired by Sir Norman Lindop. Three years later, the Lindop Committee published its report (the Lindop report), which dealt with both privacy and data protection and suggested regulatory actions including the creation of a supervisory authority. The Lindop report, despite having fostered the development of data protection in the UK, did not however end up generating any concrete, direct regulatory response until the 1984 Data Protection Act – an extensive time frame that led to the UK losing its policy advantage as a privacy and data protection forerunner in the EU.

4.1.3 Emerging national data protection legislation

The ‘70s saw the emergence of several data protection-related regulatory initiatives across Europe: the first wave of data protection legislation was enacted as a response to the rise of large-scale data processing within governments and corporations. Those initiatives were

---

699 Such as the UK, which was at the forefront of privacy and data protection development at the end of the ‘70s: See Warren and Dearnley (n 694).


701 The principles identified, which still echo in modern data protection legislation, are the following: (1) The purpose of holding data should be specified; (2) There should only be authorized access to data; (3) There should be minimum holdings of data for specified purposes; (4) Persons in statistical surveys should not be identified; (5) Subject access to data should be given, (6) There should be security precautions for data; (7) There should be security procedures for personal data; (8) Data should only be held for limited relevant periods; (9) Data should be accurate and up to date; (10) Any value judgements should be coded.

702 Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 43. Fuster reports (fn 118) that one, titled Computers and Privacy (CmND 6353), announced the government’s intention to consider legislation, while the other, titled Computers: Safeguards for Privacy (CmND 6354), dealt with computer use in the public sector. See also Warren and Dearnley (n 694).


704 The Lindop report put forth a number of principles too, which were heavily conditioned by the outcomes of the Younger committee; those principles are: (1) Data subjects should know what personal data relating to them are handled, why those data are needed, how they will be used, who will use them, for what purpose, and for how long; (2) Personal data should be handled only to the extent and for the purposes made known when they are obtained, or subsequently authorized; (3) Personal data handled should be accurate and complete, and relevant and timely for the purpose for which they are used; (4) No more personal data should be handled than are necessary for the purposes made known or authorized; (5) Data subjects should be able to verify compliance with these principles. In the interest of users; (6) Users should be able to handle personal data in the pursuit of their lawful interests or duties to the extent and for the purposes made known or authorized, without undue extra cost in money or other resources; (7) The community at large should enjoy any benefits, and be protected from any prejudice, which may flow from the handling of personal data.

705 Hondius, Emerging Data Protection in Europe (n 694); Warren and Dearnley (n 694).

706 Mayer-Schönberger (n 679) 221.
based either on secondary legislation, as happened in Hesse, Germany, Sweden, and France, or on the constitutionalisation of a right to data protection, as in Portugal, Spain, and Austria.\textsuperscript{707}

In 1970, the German Land of Hesse enacted its \textit{Hessische Datenschutzgesetz}.\textsuperscript{708} The Data Protection Act of the Land of Hesse was the first legal instrument aiming at the protection from governmental\textsuperscript{709} data processing. The \textit{Datenschutzgesetz} laid down rules on access control and data confidentiality, providing data subjects with the right to the rectification of incorrect data, and also setting up a \textit{Datenschutzbeauftragter} (Data Protection Commissioner). In 1974, Rhineland-Palatinate followed Hesse as the second Land enacting Datenschutz legislation, the \textit{Gesetz gegen missbräuchlich Datennutzung}\textsuperscript{710} (Act Against misuse of Data). In 1977, Germany enacted its first Federal Data Protection Act, the \textit{Gesetz zum Schutz vor Mißbrauch personenbezogener Daten bei der Datenverarbeitung}\textsuperscript{711} (Bundesdatenschutzgesetz, BDSG). The BDSG regulated data processing by the private sector as well, created the role of Federal Data Protection Commissioner (\textit{Bundesbeauftragter für den Datenschutz}), and introduced the notion of data protection officer (\textit{Beauftragter für den Datenschutz}). Like the \textit{Hessische Datenschutzgesetz}, the BDSG does not explicitly mention the right to privacy.

The first national law safeguarding individuals from automated data processing was however the 1973 \textit{Datalag},\textsuperscript{712} Sweden’s Data Protection Act, which regulated the conditions under which information processing was legal, and tackled issues of data security, secrecy, and accuracy.\textsuperscript{713} The \textit{Datalag} aimed at the protection of the integrity of the individuals whose data are kept in databanks held by both the public and the private sector. It explicitly referred to the concept of personal information – \textit{personuppgift} – and instituted a Data Protection Board tasked with authorising the creation and operation of machine-readable personal data registers. Other Scandinavian countries pioneered data protection legislation a few years later. Denmark followed up with the 1978 \textit{Lov om private register} (Private Registers Act) and the \textit{Lov om offentlige myndigheders register} (Public Authorities Registers Act), which regulated data banks held by the private and the public sector respectively. Also in 1978, Norway enacted its Data Registers Act as well.\textsuperscript{714}

France had been exploring the computers and privacy issue since the beginning of the ’60s, and even investigating the possibility to regulate the matter by law, without however enacting any kind of regulation for over a decade. At the beginning of the ’70s, both the \textit{Conseil d’État} and the Ministry of Justice studied the issue, and “(t)heir common conclusions described a risk inherent to computer science, from which they inferred a need to legislate”.\textsuperscript{715} In 1974, after the public expressed concern about the potential invasiveness S.A.F.A.R.I. databank project (\textit{Système Automatisé pour les Fichiers administratifs et le Répertoire des Individus}), a Commission \textit{informatique et libertés} was established at the Ministry of Justice to explore possible regulatory solutions in the area. The Commission \textit{informatique et libertés}, notably, did not try to define the freedoms (libertés) impacted by the rise of computers, framing freedom

\begin{itemize}
\item \textsuperscript{707} See Fuster, \textit{The Emergence of Personal Data Protection as a Fundamental Right of the EU} (n 19) 55 ss.
\item \textsuperscript{708} Hondius, \textit{Emerging Data Protection in Europe} (n 694) 34 ss.
\item \textsuperscript{709} The Act covered data processing in the public sector, but not in the private one.
\item \textsuperscript{710} Hondius, \textit{Emerging Data Protection in Europe} (n 694) 36.
\item \textsuperscript{711} Fuster, \textit{The Emergence of Personal Data Protection as a Fundamental Right of the EU} (n 19) 60.
\item \textsuperscript{712} Hondius, \textit{Emerging Data Protection in Europe} (n 694) 44 ss.
\item \textsuperscript{713} Mayer-Schönberger (n 679) 223.
\item \textsuperscript{714} Fuster, \textit{The Emergence of Personal Data Protection as a Fundamental Right of the EU} (n 19) 65.
\item \textsuperscript{715} Fuster, \textit{The Emergence of Personal Data Protection as a Fundamental Right of the EU} (n 19) 62.
\end{itemize}
as indivisible. Its conclusions, written by Tricot and Catala (hence the moniker “Tricot report”), were submitted to the French government in 1975. In 1976, as a consequence of the report, France adopted the loi relative à l’informatique, aux fichiers et aux libertés du 6 janvier 1978, regulating the processing of “nominal information” (informations nominatives), and setting up a regulatory authority – the Commission Nationale Informatique et Libertés (CNIL). In 1979, Luxemburg followed up with its Loi du 31 mars 1979 réglementant l’utilisation des données nominatives dans les traitements informatiques.

Aside from the use of secondary legislation to safeguard individuals from the perils related to (personal) data processing, some European countries embedded such protection in their own constitutional frameworks. Portugal regulated data processing in its 1976 Constitution’s Article 35, which ended up being a fundamental inspiration for the constitutionalisation of data protection worldwide. In 1978, Austria enacted its Bundesgesetz über den Schutz personenbezogener Daten (Datenschutzgesetz, DSG), a federal act with constitutional force which embeds the right to data protection in the Austrian constitutional milieu. The 1978 Spanish Constitution, in its Article 18(4), established a mandate to regulate the use of computers as well.

4.1.4 DATA PROTECTION’S INCEPTION AND INTERNATIONAL ORGANISATIONS

Moving from the same concerns from which the discussions on computers and privacy in the US and the emergence of (personal) data protection legislation in Europe stemmed, two major international organisations placed themselves at the vanguard of the development of a supranational data protection framework: the OECD (Organisation for Economic Co-operation and Development) and the Council of Europe. The development of the two instruments ran in parallels, and each instrument’s content is strongly linked to the other’s.

The OECD’s work on data protection started in the late ‘60s. Between 1968 and 1977, the OECD was active in the organisation of events, creation of expert groups, and publication of reports in the field of privacy and data protection vis-à-vis emerging computing technologies. In 1978, the OECD set up an Expert Group entrusted with the drafting of guidelines on the protection of privacy and the transborder flow of personal data. The OECD Guidelines were adopted in 1980, in the form of a Council Recommendation concerning Guidelines on the

---

716 Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 63.
717 Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 66.
718 Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 67; Frits W Hondius, ‘A Decade of International Data Protection’ (1983) 30 Netherlands International Law Review 103, 126.
719 Gerhard Stadler and Thomas Herzog, ‘Data Protection: International Trends and the Austrian Example’ (IIASA 1982); Hondius, ‘A Decade of International Data Protection’ (n 718) 126.
720 Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 68.
721 The OECD is an international, intergovernmental organisation, founded in 1960 on the ashes of the Organisation for European Economic Co-operation (OEEC), which aims at fostering trade and economic development.
722 Hondius, Emerging Data Protection in Europe (n 694) 57 ss.
723 See Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 76 ss.
724 The OECD’s works appear indeed to subscribe to the notion of privacy as information control put forth by Westin: see Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 76, 78.
725 Transborder data flows were a natural concern for an organisation tasked with fostering economic development.
Protection of Privacy and Transborder Flows of Personal Data,\textsuperscript{726} and subsequently modernised in 2013.\textsuperscript{727}

The OECD Guidelines apply to the processing of personal data – any information relating to an identified or identifiable individual – which “pose a danger to privacy and individual liberties”, carried out both in the private and in the public sector. They include eight basic, general principles of national application,\textsuperscript{728} and some principles of international application on the transborder flow of data and its legitimate restrictions; the Guidelines also include a part on national implementation and on international cooperation.

The Council of Europe, an international organisation founded in 1949 to support human rights and democracy in Europe, began the development of its own human rights convention from its very foundation. In 1950, its members signed the ECHR, which entered into force in 1953. Art. 8.1 ECHR safeguards the right to private and family life, home, and correspondence: contrary from the UDHR, it does not explicitly mention the right to privacy. The CoE institutions, until 1967-1968, actually avoided using the wording “privacy” within the context of Article 8.\textsuperscript{729} In 1967, the Consultative Assembly of the Council of Europe prompted its Legal Committee to adopt a resolution on human rights and technological developments, and another one on eavesdropping, wiretapping, and other privacy-intrusive technologies. In 1968, the Legal Committee presented a report to the Council of Europe’s Parliamentary Assembly, following which the Council of Europe adopted Recommendation 509 (1968)\textsuperscript{730} on Human Rights and modern Scientific and Technological Developments, which framed technical and technological progress as a potential threat to the right to privacy as framed by Art. 8.1 ECHR. Recommendation 509 also called for a study on the matter, which was carried out by the Council of Europe’s Committee of Experts on Human Rights. Following Recommendation 509,

\begin{enumerate}
\item \textsuperscript{727} The revision (Recommendation of the Council concerning Guidelines governing the Protection of Privacy and Transborder Flows of Personal Data (2013) [C(80)58/FINAL, as amended on 11 July 2013 by C(2013)79]) is the first since the original OECD Guidelines. It derived from a call by Ministers in the 2008 Seoul Declaration for the Future of the Internet Economy, answered by the OECD Working Party on Information Security and Privacy (WPISP), which agreed on the review’s Terms of Reference in 2011. The WPISP convened an Expert Group, and developed the revision of the OECD guidelines based on the Expert Group’s work. The Revision does not modify the original Guidelines’ principles, but complements them to reflect the changes in the technological and social landscape that happened during the three decades following their publication. “Two themes run through the updated Guidelines. First is a focus on the practical implementation of privacy protection through an approach grounded in risk management. Second is the need for greater efforts to address the global dimension of privacy through improved interoperability”: OECD, ‘The OECD Privacy Framework’ (2013).
\item \textsuperscript{728} Namely: Collection Limitation, Data Quality, Purpose Specification, Use Limitation, Security Safeguards, Openness, Individual Participation, and Accountability principles.
\item \textsuperscript{729} Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 82 ss.
\item \textsuperscript{730} Council of Europe, Parliamentary assembly, Recommendation 509 (1968) Human rights and modern scientific and technological developments (1968).
\end{enumerate}
the Council of Europe adopted Resolution 73 (22)\textsuperscript{731} and Resolution 74 (29),\textsuperscript{732} recommending a number of principles for the protection from unfair processing by databanks in the private and public sector respectively. In 1976, the Council of Europe set up a Committee of Experts on Data Protection,\textsuperscript{733} tasked with the preparation of a convention on data protection, to be promulgated in 1980. The preparation of the convention saw the interaction of the Council of Europe with both the OECD, which was preparing its 1980 Guidelines, and with the European Community’s institutions, which were starting to explore data protection’s regulation. In 1980, the Council of Europe published its Resolution 721 (1980),\textsuperscript{734} where it welcomed the EC institutions’ interest. At the same time, it also adopted – following a Committee Opinion and a Report on Data processing and the protection of human rights\textsuperscript{735} – Recommendation 890 (1980)\textsuperscript{736} on the protection of personal data, exploring the possibility to embed the right to personal data protection within the ECHR’s text.

The final text of the Council of Europe’s convention on data protection – by now commonly referred to as Convention 108 – was finally published in 1980, opened for signatures 1981, and entered into force in 1985. Convention 108, which mandates the enactment of national data protection legislation by the signing parties, generated a new wave of national data protection laws. An Additional Protocol,\textsuperscript{737} containing additional norms on supervisory authorities and on transborder data flows, was open to signature in 2001. In February 2016, the Council of Europe, on the basis of Convention 108, also began drafting guidelines on the protection of personal data in the big data context. The “Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data” (hereinafter Guidelines) were finally adopted in January 2017.\textsuperscript{738}

Convention 108 targets data protection, “i.e. the legal protection of individuals with regard to automatic processing of personal information relating to them”,\textsuperscript{739} moving the syntagm “data protection” away from the eminently German Datenschutz, and enshrining it in a binding international regulatory instrument. It connects data protection with the safeguard of rights and freedoms in general, but links it with the right to privacy as hallowed by Article 8 of the ECHR in particular. It comprises general provisions on the Convention’s scope and mechanics, high-level data protection principles, and norms on transborder data flows and on mutual

\textsuperscript{731} Council of Europe, Committee of Ministers Resolution (73) 22 On the Protection of The Privacy of Individuals Vis-A-Vis Electronic Data Banks in The Private Sector (1973). Resolution 73 (22) can indeed be considered the first instrument regulating personal data protection at an international level: Hondius, ‘A Decade of International Data Protection’ (n 718) 103.

\textsuperscript{732} Council of Europe, Committee of Ministers Resolution (74) 29 On the Protection of The Privacy of Individuals Vis-A-Vis Electronic Data Banks in The Public Sector (1974).

\textsuperscript{733} Hondius, ‘Data Law in Europe’ (n 697) 105.


\textsuperscript{735} CoE Committee on Science and Technology, Report: Data processing and the protection of human rights Doc. 4472 (1979); CoE Committee on Legal Affairs and Human Rights, Report: Data processing and the protection of human rights Doc. 4472 (1979).


\textsuperscript{737} Council of Europe, Additional Protocol to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, regarding supervisory authorities and transborder data flows, ETS 108 (2001).


cooperation and assistance. Thus, despite being formally directed at safeguarding rights and freedoms, privacy in particular, through the concept of personal data protection, it is also substantially concerned with ensuring the free flow of information. Convention 108 is also explicitly focused on automated processing, rather than on data processing *tout court*.

The OECD and the Council of Europe’s initial approach on data protection has been historically entwined, their development running almost in parallels, and influencing the European Community’s developments on the matter. They both highlighted the transborder dimension of (data) privacy and data protection, and Convention 108 in particular reinforced the notion of data protection as separate from – albeit connected with – the right to privacy.

**4.1.5 A EUROPEAN DATA PROTECTION FRAMEWORK**

The EU is arguably the main regulatory actor within the worldwide data protection landscape nowadays. European data protection legislation’s territorial scope is quite expansive, and its reach is not limited to what it explicitly covers: EU data protection had (and still has) a stark influence on the legislation of several non-EU countries. At the time of writing, the existing secondary legal framework is being reformed, and its drafting reflects the aftermath of the recent (quasi)constitutionalisation of the right to personal data protection by the Lisbon treaty. This subsection, and the following ones, account for the rise of the EU right to data protection as linked with the national and international developments above, for its (quasi)constitutionalisation within the EU’s fundamental rights framework, and for the direction taken by its secondary legislation’s reform process.

The beginning of the EC’s involvement in the area of data protection did not however stem exclusively from concerns about Europeans’ fundamental rights and freedoms, but also from economic reasons, particularly from concerns about International Business Machine’s (IBM) market dominance. In this context, in 1973 the Commission published a “Community policy on data processing” in the form of a Communication to the Council, in which it suggested to foster the development of the data processing industry, and incidentally noted that doing so would have required ensuring an appropriate level of protection for individuals. In 1974, the issues raised by the Commission in its Communication started being explored by the European

---

740 Bennett and Raab (n 239) 90.
741 The development of EU data protection law paralleled the development of the EU itself, which appears thus useful to summarise here. Roughly, the European Coal and Steel Community, created by treaty in 1951, gave way to the European Economic Community (EEC), formed through the Rome Treaty of 1957. The 30 years after the Rome Treaty saw relatively little change in the EEC, aside from the institutional changes introduced by the Single European Act (SEA). The situation changed in 1992, with the signature of the Treaty on European Union (TEU) in Maastricht, which created the EU and renamed the EEC into EC (the European Community), introducing several institutional and substantive changes to the Treaty of Rome. The process continued with the Treaty of Amsterdam of 1997, applicable from 1999 on, and with the treaty of Nice in 2000. The new millennium started with discussions about a Constitutional Treaty, the progress of which was halted after France and the Netherlands rejected it after a national referendum. The instances underlying the failed Constitutional Treaty have however been at least partly advanced within the Lisbon Treaty of 2009, through which the EU replaced and succeeded the EC, and that changed and amended both the TEU and the EC Treaty (now Treaty on the Functioning of the European Union, or TFEU). See *inter alia* Paul Craig, ‘Development of the EU’ in Catherine Barnard and Steve Peers (eds), *European Union Law* (Oxford University Press 2014).
742 See e.g. Birnhack (n 66); Anu Bradford, ‘The Brussels Effect’ (2012) 107 Nw. UL Rev. 1; Greenleaf (n 66).
743 The ePrivacy Directive is currently on its way to be substituted with an ePrivacy Regulation.
Parliament as well, and at the end of the year the Legal Affairs Committee of the European Parliament (with Lord Mansfield as rapporteur) received mandate to prepare a report on the protection of the rights of the individual in the face of the growing technical progress in data processing.

Between 1975 and 1976, as a consequence of the Mansfield report, the European Parliament adopted two Resolutions “on the protection of the rights of the individual in the face of developing technical progress in the field of automatic data processing”, urging for legislative action on data processing and individual freedom. The second Resolution led to the drafting of a report by the Data Processing and Individual Rights Sub-committee, with Mr. Bayerl as rapporteur, which explored the developments in the field of human rights and data processing within the EC member states, the Council of Europe, and the OECD. In 1979, the European Parliament adopted – following the Bayerl report – another Resolution “on the protection of the rights of the individual in the face of technical developments in data processing”, putting forth a series of recommendations meant to form the basis for the future Community regulations on data processing.

In the following years, both the OECD Guidelines and Convention 108 saw the light of day. The Commission, in 1981, published a Communication urging Member States to ratify Convention 108. In 1982, the European Parliament adopted another Resolution “on the protection of the rights of the individual in the face of technical developments in data processing”, where it welcomed Convention 108 and solicited its ratification by the EC Member States. Between 1985 and 1990, the development of the Schengen Agreement and of the relative implementing Convention, which established the Schengen Information System (SIS) and explicitly dealt with personal data protection within the latter’s operation, contributed in fostering the need for a pan-European data protection framework.

Finally, in 1990, the European Commission published a Communication proposing a regulatory package disciplining the processing of personal information. Aside from the harmonisation sought by the regulatory proposal, the legislature’s aim was to fix a situation that was deemed “generally speaking [...] unsatisfactory within the Community”, leading to an inadequate level of legal protection of the European citizen. The reasons for the adoption of the regulatory package were largely due to the emergence of the Single Market, to be completed by 1993, and to the rise in importance of the protection of human rights and freedoms within the EC. The legislative package’s content was massive: a Proposal for a Directive on the

---


746 Hondius, ‘A Decade of International Data Protection’ (n 718) 105.


750 See Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19) 123 ss.


752 Blume (n 751).

753 Blume (n 751).
The initial development of a pan-European data protection framework, largely fostered by the works of the OECD and of the Council of Europe between the ‘70s and the ‘80s, was refined largely in the following two decades, culminating in the Commission’s 1990 Proposal. European data protection within the EC (and then the EU) began at a fundamental turning point in computing history: the shift from the mainframe era in the ‘60s and ‘70s – the advent of which raised the original “computers and privacy” issue – to the Personal Computer one, and the beginning of the diffusion of the (commercial) Internet and of the World Wide Web.

Data protection adapted to technology. The period running from the ‘60s to the ‘80s saw private actors getting close to public ones in their information gathering and processing capabilities, partly moving the general concern away from the State and towards the modern corporation. From the ‘80s onwards, computers became personal – the availability of computation to individual consumers at least partly democratised processing power. In 1991, the general population was also introduced to the World Wide Web, and thus to the Internet, whose development had been under way by roughly two decades.

The problems deriving from the enhanced power and information asymmetry brought forth by the mainframe computer and by centralised databanks were however not solved nor mitigated by computing capacity becoming affordable at household level, nor by the decentralisation on which the Internet has always been inspired. The general rise in the population’s computing capacity ran in parallel with a much starker growth in the capabilities available to both the public sector and the corporate milieu.

The time span running between the framing of privacy and data protection by the OECD Guidelines and by the Council of Europe’s Convention 108, and the 1990 Commission’s proposal, saw a powerful change in computing technologies – the very technologies from which the original “computers and privacy” issue stemmed. The decade between 1990 and 2000, which was increasingly characterised by the depth with which the Internet and the so-called Web 1.0 shaped society, saw the drafting and enactment of the EU data protection framework – the Data Protection Directive (DPD) in particular. At the same time, despite being crafted well within the PC era and at the beginning of the World Wide Web’s one, the

754 Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19); Blume (n 751).
original European data protection framework was heavily reliant on the principles developed during the mainframe age, especially as expressed in the OECD Guidelines and Convention 108. The DPD, since its drafting, was thus already lagging behind the state of the art, due to the rapid pace of innovation in the computing and networking technologies from which the original “computers and privacy” issue arose.

4.1.5.2 Developing EU Data Protection Law

The Commission’s Proposal for a Directive on the protection of individuals in relation to the processing of personal data was slated to become the cornerstone of the European data protection framework, and so it did, eventually becoming, five years later, the DPD. The 1990 proposal was subject to the opinion of the Economic and Social Committee,\(^757\) which was delivered in 1991, and to the European Parliament’s one,\(^758\) who in 1992 proposed many changes in the original text. In 1992, the Commission published a modified proposal.\(^759\) In the meantime, the enactment of the Treaty of Maastricht – which was signed in 1992 and entered into force in 1993 – modified the applicable legislative procedure, and the European Parliament thus gained co-decision competences. At the Council level, discussions amongst Member States saw the divergence between two blocks of States who pushed for and against the adoption of the Directive, respectively, which held up the Directive’s development for a couple of years. The Council reached a common position in 1995;\(^760\) a few months later, the Data Protection Directive was approved, and the European Community finally had its own data protection instrument, soon to be transposed in national law.


and the protection of privacy in the electronic communications sector (the ePrivacy Directive), which was amended in 2009, and is currently in the process of being reformed. Notably, Article 14 of the ePrivacy Directive allowed member states to enact, for specific purposes relating e.g. to policing, security, or defence, legislative measures mandating the retention of data for a limited period of time. National legislatures, as a consequence, implemented their own laws requiring data retention. The EC, following the attacks in Madrid in 2004 and in London in 2005, purportedly with the aim of harmonising the national legislation on data retention, initially proposed the adoption of a Framework Decision, framing the issue as a Third Pillar matter. Following opposition to the text from some Member States, the proposal for a data retention instrument resurfaced as a draft Directive, which was eventually approved as Directive 2006/24/EC on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks. The Data Retention Directive was declared invalid by the ECJ in 2014 but to this day many national legislations implementing it still linger on.

The 1990 Commission’s Communication contained also a pledge, by the European Commission, to conform the European Community’s institutions to the same level of protection granted to data subjects by the DP Directive. In 1997, the Treaty of Amsterdam introduced Article 286 in the Treaty establishing the European Community (EC Treaty), which holds that EC acts on data protection must apply to EC institutions and bodies as well. In 1999, the Commission proposed a Regulation on the protection of personal data by Community institutions and bodies, which eventually lead to Regulation (EC) No 45/2001 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data.

---


764 See the Commission’s Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications), COM/2017/010 final - 2017/03 (COD) and its other versions as resulting from the ordinary legislative procedure.


766 CJEU, Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others, Joined Cases C-293/12 and C-594/12, 8 April 2014, ECLI:EU:C:2014:238.


The intent of the 1990 Commission’s Proposal was also to extend its principles to processing activities not governed by EC law.\textsuperscript{769} to this purpose, the Proposal contained a Draft resolution, which did not however generate any concrete reaction. The reticence in adopting data protection instruments outside of the first pillar’s domain lasted for fifteen years, until the Commission, in 2005, proposed Council Framework Decision on the protection of personal data processed in the framework of police and judicial cooperation in criminal matters\textsuperscript{770}. Framework Decision 2008/977/JHA,\textsuperscript{771} which derives from the Commission’s proposal, was adopted by the Council three years later, and has now been repealed by the Police Directive during the data protection reform process started in 2009.\textsuperscript{772}

4.1.5.3 THE CHARTER AND DATA PROTECTION’S (QUASI)CONSTITUTIONALISATION

As the review above has shown, the history of the European right to personal data protection has been historically entwined with the right to privacy, particularly in relation to the development of computing technologies. Such connection is testified, \textit{inter alia}, by the DPD’s Article 1(1), which clarifies that the directive is meant to safeguard “the fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the processing of personal data”.\textsuperscript{773} From a formal perspective, the admitted blurriness between the right to privacy and the one to data protection also derives from the fact that the ECHR – the human rights catalogue Europe used to rely on – safeguarded both rights through the same Art. 8. Incidentally, accession to the ECHR was one of the paths available to the EU to adopt its own catalogue of fundamental rights, aside from drafting its own one.\textsuperscript{774}

In 2000, however, after a long and tortuous path,\textsuperscript{775} the EU enacted its Charter of Fundamental Rights\textsuperscript{776} (the Charter): a list of rights, common to the European constitutional tradition, meant to have a prominent status within the EU. The drafting of the Charter was entrusted by the European Council to an organ – the Convention – which worked on the text for roughly two years, from 1999 to 2000. it seemed clear from the outset that the right to data protection was likely to be in the catalogue, and as a separate right from the one to private life. The Charter

\textsuperscript{769} I.e. outside what was known as the “first pillar” of the EU: which corresponded to the former EC. The second pillar was the Common Foreign and Security Policy (CFSP), while the third pillar encompassed a broad range of issues, from immigration to police and judicial cooperation in criminal matters.


\textsuperscript{772} See section 4.3.6 below.

\textsuperscript{773} The debate about the EU’s accession to the ECHR has been ultimately solved through the Lisbon treaty, which modifies the TEU (formerly TEC) by providing both the competence and the obligation to do so (See Art. 6.2 TEU): Eleanor Spaventa, ‘Fundamental Rights in the European Union’ in Catherine Barnard and Steve Peers (eds), \textit{European Union Law} (Oxford University Press 2014) 235.

\textsuperscript{774} Discussions about the EU adopting a fundamental rights system were being held since the end of the ’70s, when EU institutions started debating whether to adopt the ECHR or to draft the EU’s own fundamental rights catalogue.

was officially adopted on the 7th of December 2000 by the European Parliament, the Council, and the Commission, its legal force still unclear. Indeed, while solemnly proclaimed by the European institutions, the Charter was not legally binding at its beginning. It remained void of binding force for almost a decade, and the ECJ itself started referring to the Charter only after quite some time from its proclamation.

The Lisbon Treaty was signed in 2007 by the EU Member States, and entered into force in 2009. It modified both the Treaty on the European Union (TEU) and the Treaty on the European Community (TEC), the latter of which became the Treaty on the Functioning of the European Union (TFEU). Article 6 of the TEU, as reformed by the Lisbon Treaty, thus became the main provisions regarding EU fundamental rights, describing how they come to be hallowed in the EU legal order; its first paragraph recognizes for the charter the same legal force than the EU treaties (TEC, now TEU, and TFEU). From 2009 onwards, the Charter thus acquired binding force, and the EU – while still waiting for the formal accession to the ECHR – had its own fundamental rights catalogue.

The Charter – particularly after the Lisbon Treaty made it binding, and the ECJ started referring to it in its judgements – is arguably the closest instrument to a constitution within the EU legal framework. The complex interplay between EU primary and secondary law on one hand and Member States constitutional provisions on the other, and the failure of the enactment of a European Constitution over the past two decades, makes it not possible to define the Charter as a proper constitutional system. Its prominence within the EU’s legal framework is however unquestioned, and this dissertation will therefore refer to the Charter, the EU Treaties, and the ECHR, as the EU’s (quasi)constitutional system.777

The enactment of the Charter had a stark effect on the recognition of data protection as a separate fundamental right from the right to privacy. Indeed, differently from the ECHR, the Charter contains two different articles differentiating the right to private life from the one to the protection of personal data, Articles 7 and 8 respectively. Article 7, titled “Respect for private and family life”, recites that “Everyone has the right to respect for his or her private and family life, home and communications”. Article 8, “Protection of personal data”, hallows the right to the protection of personal data,778 which must be processed fairly, for specified purposes, and on the basis of the consent of the person concerned or some other legitimate ground. Article 8 also enshrines the right of access one’s personal data, and to have it rectified.779 Compliance must be under the control of an independent supervisory authority.780Limitations to the rights to privacy and data protection, as for all the other rights in the Charter, must be in accordance with its Article 52: “Any limitation on the exercise of the rights and freedoms recognised by this Charter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made

---

777 For a short primer on European constitutionalism (i.e. the conceptual framework identifying what a constitution is or should be) see Robert Schütze, ‘Constitutionalism and the European Union’ in Catherine Barnard and Steve Peers (eds), European Union Law (Oxford University Press 2014). The Constitutional status of EU law is often claimed in doctrine, sometimes also by the ECJ: Bruno de Witte, ‘EU Law: Is It International Law?’ in Catherine Barnard and Steve Peers (eds), European Union Law (Oxford University Press 2014) 190.

778 Charter, Art. 8(1): “Everyone has the right to the protection of personal data concerning him or her”.

779 Charter, Art. 8(2): “Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.”

780 Charter, Art. 8(3): “Compliance with these rules shall be subject to control by an independent authority”.
only if they are necessary and genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others”. 781

4.1.5.4 The Lisbon Treaty and Art. 16 TFEU

The Lisbon Treaty did not only give binding force to the Charter, but abolished the pillar structure on which the EU was based too, and introduced also a new provision on data protection – Art. 16 TFEU – replacing the former Art. 286 EC. 782 Art. 16 TFEU’s wording has a broader scope than its predecessor, 783 and reads as follows:

“Everyone has the right to the protection of personal data concerning them.

The European Parliament and the Council, acting in accordance with the ordinary legislative procedure, shall lay down the rules relating to the protection of individuals with regard to the processing of personal data by Union institutions, bodies, offices and agencies, and by the Member States when carrying out activities which fall within the scope of Union law, and the rules relating to the free movement of such data. Compliance with these rules shall be subject to the control of independent authorities.

The rules adopted on the basis of this Article shall be without prejudice to the specific rules laid down in Article 39 of the Treaty on European Union”.

The introduction of Art. 16 is an important change for European data protection. Art. 286 EC was a first-pillar instrument only, and thus covered only the areas formerly encompassed by EC law; Art. 16 TFEU, on the contrary, has general application, and thus covers all areas of EC law. 784 Its position within the Treaties’ structure is more prominent than the position held by the former Art. 286 EC, which it replaced, signalling the growing importance of personal data protection within the EU’s legal milieu. The materialisation of Art. 16 restates the existence of personal data protection as a separate right within the EU legal order. 785

Perhaps most importantly, Art. 16 TFEU allows to overcome the issues deriving from the usage of Art. 114 TFEU (formerly Art. 95 TEC) as a legal basis for the adoption of the DPD. 786 One of the main narratives surrounding the emergence of the European Union portrays it as an

781 Charter, Art. 52(1).
783 Article 286 of the EC Treaty read as follows:
“1. From 1 January 1999, Community acts on the protection of individuals with regard to the processing of personal data and the free movement of such data shall apply to the institutions and bodies set up by, or on the basis of, this Treaty.
2. Before the date referred to in paragraph 1, the Council, acting in accordance with the procedure referred to in Article 251, shall establish an independent supervisory body responsible for monitoring the application of such Community acts to Community institutions and bodies and shall adopt any other relevant provisions as appropriate”.
785 Fuster, ‘Fighting for Your Right to What Exactly - the Convoluted Case Law of the EU Court of Justice on Privacy and/or Personal Data Protection’ (n 649) 266.
786 The DPD’s dual objective of ensuring the free flow of data and protecting individuals’ rights has been deemed as problematic e.g. in Orla Lynskey, ‘From Market-Making Tool to Fundamental Right: The Role of the Court of Justice in Data Protection’s Identity Crisis’ in Serge Gutwirth and others (eds), European Data Protection: Coming of Age (Springer Netherlands 2013).
economic project that still carries the market-making rationale of the seminal European Economic Community (EEC).

Yet, as it has been noted,787 the decades where the EU (and the EC, and the EEC before that) functioned without its own fundamental rights catalogue were the result of political pragmatism, rather than of an eminently economic focus. The delays in adopting fundamental rights legislation at a primary level were mostly driven by the strategic and political drivers of state actors, which were reticent to hinder the integration process by adding an additional fundamental rights dimension to the political processes that were underway,788 particularly in light of a series of failures in modifying the institutional asset of the EU (and of its predecessors), such as the abandonment of the European Defence Community and of the European Political Community, and more recently of the Treaty establishing a Constitution for Europe. EU legislatures, arguably, always had fundamental rights in mind, but have always been limited in their action by the principle of conferral,789 and by the need for a suitable legal basis justifying the enactment of new legislation.

The DPD represents the perfect example of how a fundamental rights instrument had to be stuffed inside a market-shaped trojan horse. The DPD’s legal basis was Art. 114 TFEU, which allows the EU legislature to adopt measures “which have as their object the establishment and functioning of the internal market”: a market-making tool. On one hand, that is coherent with one of its two objectives, namely ensuring the free flow of data. On the other hand, the fundamental rights dimension of the DPD has been unambiguously prevalent over its economic underpinning, both considering its substantive provisions and with regard to its interpretation in the CJEU’s case law.790

In consideration of the weight of the DPD’s fundamental rights’ dimension, and particularly in comparison with its market-making rationale, the suitability of Art. 114 TFEU as a legal basis might be somewhat questionable.791 The introduction of Art. 16 TFEU within the EU’s (quasi)constitutional framework allowed the General Data Protection Regulation (GDPR) to avoid this conundrum: Art. 16 TFEU now provides for a separate legal basis justifying EU data protection legislation, and there is no longer the need to stretch the market-oriented Art. 114 TFEU to promote data protection’s fundamental right facet.792 At present times, after the Charter’s publication – its legal strength cemented by the Lisbon Treaty – and the introduction of Art. 16 TFEU, it appears difficult to argue against the conceptual and formal separation between the right to privacy and the one to data protection within the EU legal order.

788 Janciute (n 787) 5.
789 Art. 5 TEU.
790 See Lynskey, ‘From Market-Making Tool to Fundamental Right: The Role of the Court of Justice in Data Protection’s Identity Crisis’ (n 786), and the cases cited therein.
791 The same considerations go, mutatis mutandis, for the explicit introduction of the wording ‘open data’ in the latest versions of the recast of the PSI Directive (see the previous chapter). To what extent is Art. 114 TFEU a suitable legal basis for the Recast of the PSI Directive if, to the economic rationale that underpinned EU PSI legislation from its very beginning, one adds to it the whole spectrum of rights and freedoms underlying open data?
792 Which is, incidentally, even more beneficial in areas where the market-making rationale of data protection is non-existent, such as justice and home affairs, or police activities – the areas formerly encompassed by the c.d. second and third pillars: see Hijmans and Scirocco (n 784).
4.1.5.5 THE EU DATA PROTECTION REFORM

After almost 20 years from the inception of the DPD, the European institutions started working on updating a regulatory system that was admittedly getting outdated, and where the fragmentation due to differences deriving from the Member States’ different implementations of the DPD were deemed to be hampering the development of the Single Market.793 The works for the reform started in 2009,794 with the setting up of two public consultations (held from 2009 to 2011) and the release of a Commission’s Communication on “A comprehensive approach on personal data protection in the European Union”795 in 2010. The Consultation of the European Commission on the legal framework for the fundamental right to protection of personal data and the following Communication were followed by the publication of the views of, inter alia, the Parliament796, and the Article 29 Data Protection Working Party (A29WP).797

In 2012, the Commission Published its “Safeguarding Privacy in a Connected World - A European Data Protection Framework for the 21st Century”,798 officially starting the EU data protection reform process. The reform aimed at repealing both the DPD and Framework Decision 2008/977/JHA, to be renewed through the Commission’s proposals for a General Data Protection Regulation and for a Directive on the protection of individuals with regard to the processing of personal data by competent authorities799 (the Police Directive) respectively.800

The Commission’s proposal for a GDPR – which took a form of a Regulation, rather than a Directive, to enhance the future law’s harmonising effect801 – set out to reform the DPD, adapting EU data protection to twenty years of technological progress and globalisation, and overcoming the differences arising from the DPD’s implementation in national law. It aimed at

794 Instrumental to this effect appears to have been the conference titled “Personal data - more use, more protection?”, organised in May 2009 by the Commission: see http://ec.europa.eu/justice/newsroom/data-protection/events/090519_en.htm, last access June 2017.
799 Proposal for a Directive of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data (2012) COM/2012/010 final - 2012/0010 (COD).
800 Processing for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties is however out of the scope of this dissertation, and this section will thus focus on the GDPR only. The proposal for a Police Directive eventually lead to Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Framework Decision 2008/977/JHA (2016) OJ L 119/89.
doing so both to enhance individuals’ fundamental rights protection, and to boost the Single Market. The Commission’s Proposal for a GDPR maintained a large degree of continuity802 with the principles of European data protection developed during the previous decades, and thus with the Fair Information Processing Principles doctrine. The Commission’s Proposal had an initially positive reception.803

The Parliament’s first reading,804 which was adopted in March 2014, significantly amended the original Commission’s Proposal. Following the ordinary legislative procedure,805 the Council reached a general approach806 Monday 15 June 2015, signalling the start of the so-called ‘trilogue’ phase – an institutional dialogue between the EU Commission, Parliament, and Council. The political agreement between Parliament and Council was reached on the 15th of December 2015.807 The 8th of April 2016 the Council adopted its position at first reading,808 which was then approved by the Parliament at second reading in April 2016.809 The GDPR was then signed on the 27th of April 2016, and published in the Official Journal of the EU on the 4th of May810. It has been applicable since the 25th of May 2018.811

The EU data protection reform process did not end with the adoption of the GDPR and the Police Directive. On the 10th of January 2016, the Commission issued a Proposal for a


805 Previously known as ‘codecision’, it is the main legislative instrument of the EU, regulated by Art. 294 of the TFEU. In the ordinary legislative procedure, legislative acts start from a Commission proposal, and require the agreement of the Parliament and the Council on a specific text. Such agreement can be reached either at first reading, if the Council approves the Parliament’s first reading position; at second reading, if each institution approves the other’s amendments on second reading; or if – after convening a conciliation committee – a joint text is approved on third reading. If a proposal is rejected at any stage of the procedure, or a compromise cannot be reached after the conciliation procedure, the proposal is not adopted, and that procedure ends.

806 Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) - Preparation of a general approach (2015) 2012/0011 (COD).


811 See GDPR, Art. 99.
Regulation concerning the respect for private life and the protection of personal data in electronic communications, the ePrivacy regulation. The proposal is meant to repeal Directive 2002/58/EC – the ePrivacy Directive – substituting it with an instrument in line with the GDPR and with the ever-changing technological landscape.

4.1.5.6 The Ubiquitous Computing Era

The right to personal data protection has been largely driven by the social disruptions deriving from a rapidly mutating technological landscape, and from its growing importance within every facet of modern life. The origin of data protection, from the US to the EU, is tied to the birth and growth of mainframe computing and centralised databanks. The development of the original European data protection framework has been driven by the rise of personal computing and of the Internet in the ‘90s; the EU data protection reform, which started in 2009, has a completely different technological context.

Computing has, for starters, become ubiquitous. Portable computing devices grant constant wireless access to the Internet, through a combination of public Wi-Fi and carrier data. Smartphones are omnipresent, and often tied, at a software level, to a platform ecosystem allowing external developers to provide third party applications. Personal computing has become truly personal – from desktops tied to the whole household to a multiplicity of portable computing devices linked to a single, individual user. The notion of computing device has, itself, been expanded by what is by now commonly referred to as the Internet of Things (IoT) – the digitalisation, instrumentation, and networking of all kinds of objects that were not originally designed to be connected to the Internet.

Aside from single objects, entire environments of increasing scale are being digitised, instrumented, and networked. From individual buildings to entire streets and neighbourhoods, a plethora of “living lab” and “smart city” initiatives are expanding the scale and scope of what computing and data processing mean nowadays. At an even broader level, cloud computing\(^8\) technologies have freed information from the constraints deriving from local storage and processing: the cloud,\(^8\) despite resting on the physicality of servers and data centres, has factually added an additional territory,\(^8\) another global layer on which human activities are performed. More recently, due to the fact that cloud computing is a paradigm that fosters centralisation, and hence information and power asymmetries, decentralised alternatives (e.g. through blockchain technologies)\(^8\) are also being actively sought.

Software itself has changed, becoming responsive, adaptive, introducing questions about its own agency\(^8\) and the responsibility it generates. At the dawn of the fourth industrial revolution\(^8\), advances in Artificial Intelligence (AI) and Machine Learning (ML) are introducing a set of brand-new problematic\(^8\) bound to challenge current regulatory frameworks. Big

---

8\(^1\)Mell and Grance (n 109).
8\(^2\)From a legal perspective on cloud computing, see Millard (n 346).
8\(^3\)See Bratton (n 646); Benjamin H Bratton, ‘The Black Stack’ (2014) 53 E-flux Journal.
8\(^5\)Hildebrandt (n 288).
8\(^7\)See e.g. in Dimitra Kamarinou, Christopher Millard and Jatinder Singh, ‘Machine Learning with Personal Data’ [2016] Queen Mary School of Law Legal Studies Research Paper No. 247/2016; Lilian Edwards and Michael Veale,
data\textsuperscript{819} – information sets too big and dynamic to be processed with customary processing methods – is already revolutionising all sectors traditionally based on ("small") data processing, from healthcare to transportation to social research. The World Wide Web itself – arguably the prime and most extraordinary application of the Internet – is shifting towards a Semantic Web,\textsuperscript{820} a radical expansion of the interconnectedness characterising the resources on the Web, aiming at promoting common data formats and protocols, to foster data sharing and re-use.

The EU data protection reform is happening in times of rapid change, fluctuations that are eminently driven by technological progress alone. Technology however is inherently political, and has a stark regulatory capacity. The interplay between regulation by law and regulation through technology is bound to discipline personal data processing within and without the EU. It appears hardly possible to forecast the result of this interplay – the novelties introduced by the EU data protection reform are still to be applied, and the impact of current technological development is yet to be seen – but there are some factors that arguably remained constant across the development of the right to data protection, and that are expected to continue being present after the restructuring.

Technological development impacted society radically in the past decades. Computing became ubiquitous and pervasive, growing exponentially in scale and scope. An increasing number of activities and functions is being automated, digitised, and networked, leading to a growing reliance on technology in everyday endeavours. Technology also enabled new and innovative business models, and changed the workings of many existing ones. The platform economy – the Facebooks and the Ubers, where several actors interact within “multisided digital frameworks that shape the terms on which participants interact with one another”\textsuperscript{821} – is shaping goods and services provisioning in ways that are yet to be fully understood. The world transformed a lot since the “personal computer era”, and even more so since the mainframe one. The right to personal data protection has been evolving to keep pace with such change.

However, most of the issues which personal data protection was developed to tackle have arguably remained a constant over the years. True computing and informational powers remain largely centralised: aside from the data gathered by the public sector,\textsuperscript{822} a (larger) array of private actors\textsuperscript{823} still concentrate overwhelming amounts of information and processing power, be it through cloud computing rather than through mainframes. There is still a (growing) power and information asymmetry running between individual data subjects and data controllers, and the set of rights that can be potentially dented by unfair information processing is still the same. The essence of the need for control over one’s personal information – informational self-determination – remained largely unaffected by technological progress, albeit such need has certainly grown. Data protection continues to be a highly


\textsuperscript{820} See Mayer-Schönberger and Cukier (n 13); Kitchin, \textit{The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences} (n 4).


\textsuperscript{822} The public sector – national and local administrations – have traditionally been amongst the entities that have the highest availability of information, both of a personal and a non-personal nature.

\textsuperscript{823} See Christl and Spiekermann (n 254); Christl, Kopp and Riechert (n 254).
adversarial regulatory setting, where the interests of the data controller oftentimes clash with the ones of the data subjects.

The passing of time has shown that the hypothesis for which everything that can be automated will be automated, everything that can be datafied will be datafied, and everything that can be used for surveillance and control will be used for surveillance and control, in the absence of a countervailing push.\textsuperscript{824} The right to personal data protection has come to embody such a countervailing drive, explicating itself in a set of informational rights and obligations that go beyond the right to privacy, and that are specifically aimed at curbing the power asymmetries deriving from information processing, rather than at the protection of private and family life, home, and communications (however broadly constructed).

4.2 Distinguishing the Rights to Privacy and Data Protection

The preceding sections delineated the development of the right to personal data protection in the EU, and its formal differentiation from the right to privacy. Data protection has been framed as a set of rules aiming at disciplining personal information processing in order to protect individuals from unjustified processing of their personal details\textsuperscript{825}. The distinction between data protection and privacy, due to their close connection and to the latter’s \textit{sui generis} development, is however admittedly still somewhat blurry\textsuperscript{826}. This section further clarifies the substantial distinction between the two rights.

Privacy and data protection are historically and conceptually entwined. Before the Charter, at a primary level, data protection was based on an extensive interpretation of Art. 8 ECHR, read as enshrining both the right to privacy and the one to data protection, the latter particularly after Convention 108. With the Charter, the EU got its own fundamental rights catalogue, and the distinction between privacy and data protection has finally been formalised.

At a substantive level, such distinction is however much less clear, and the two rights are still often conflated. Still, to pinpoint the essence and rationale of the right to data protection, moving from what makes it distinct from the right to privacy seems a sound starting point. In consideration of the multi-level structure of fundamental rights\textsuperscript{827} in Europe, this section frames the right to privacy as enshrined in the Charter, in the ECHR, and in the ECtHR’s jurisprudence,\textsuperscript{828} and the right to data protection as deriving from the Charter, the CJEU’s and the ECtHR’s jurisprudence, but also from the legislative framework implementing it in secondary legislation. The development of data protection is particularly entwined with the secondary legislation that preceded its (quasi)constitutionalisation – the DPD in particular – and it can be argued that the introduction of data protection as a fundamental right drew legitimacy from and gave legitimacy to the pre-existing secondary law.

\textsuperscript{824} See Zuboff, ‘The Surveillance Paradigm: Be the Friction - Our Response to the New Lords of the Ring’ (n 234).

\textsuperscript{825} De Hert and Gutwirth (n 236) 4.

\textsuperscript{826} See Fuster, ‘Fighting for Your Right to What Exactly - the Convoluted Case Law of the EU Court of Justice on Privacy and/or Personal Data Protection’ (n 649).

\textsuperscript{827} See Fabbrini (n 298).

\textsuperscript{828} Member States’ constitutional traditions differ between each other when it comes to conceptualizing privacy, and the CJEU and ECtHR’s jurisprudence provides for the degree of harmony and uniformity that analysing individual Member States’ constitutional jurisprudence would not provide.
It is possible to distinguish between three different doctrinal constructions of the relationship between the rights to privacy and data protection, all of which can be reasonably justified through literature and jurisprudence. The first model sees privacy and data protection as complementary rights, concurrently aiming at the protection of human dignity. The second one frames personal data protection as a facet of the right to privacy. The third model sees data protection as a separate right, covering a range of interests which is different from — albeit partly overlapping with — the right to privacy.

This dissertation moves from the third model, and thus frames data protection as a standalone right, meant to safeguard a large array of underlying interests, and that partly overlaps with the right to privacy. This position is grounded in the formal distinction between the two rights as resulting from the Charter and as confirmed by the Lisbon Treaty, and in the substantial differences running between the two rights. As for the latter, there are four main dissimilarities to be considered when framing the substantial distinction between privacy and data protection: the differences in the horizontal applicability between the two rights; the differences in their material scope; the fact that they cover different — albeit overlapping — categories of information and activities; and that data protection grants a wider set of informational rights than privacy. Further insights into the nature of data protection can also be drawn by analysing the structure of Art. 7, Art. 8, and Art. 51 of the Charter as compared with Art. 8 ECHR. Those points are expanded in the subsections that follow.

4.2.1 Horizontal and vertical effect: Does it matter?

As Lynskey notes, it might be useful to consider the overlap between the right to privacy and Art. 8 ECHR, on the one hand, and the EU data protection regime, on the other. Can insights about the different substance of the two rights be distilled from the differences in the subjects they mean to bind?

Lynskey holds that a first distinction could derive from the applicability of each framework to private relations. While Art. 8 ECHR (and Art. 7 CFR for that matter) do not have direct horizontal effect, the fundamental right to data protection ex Art. 8 CFR is implemented through secondary legislation that is applicable to public and private parties alike, and has thus direct horizontal effect. From this perspective, the applicability of data protection seems to be broader than privacy. However, Lynskey also notes that ECHR rights can have an indirect horizontal effect by virtue of the positive State duty doctrine, which obliges states to actively safeguard ECHR rights. Moreover, ECHR rights can be (and often are) taken in consideration by national courts in their judgements, akin to how German constitutional doctrine constructs the indirect horizontal effect (mittelbare Drittwirkung) of constitutional provisions. Finally, EU law aside, ECHR rights have been transposed into national legislation, or were enshrined in the states’ national law even before the ECHR. In a nutshell, while EU data protection has a direct effect between private parties, the ECHR privacy regime has an eminently vertical effect. The distinction is however mitigated by a number of factors (positive state duties,

829 Lynskey, The Foundations of EU Data Protection Law (n 19) 91 ss.
830 Lynskey, The Foundations of EU Data Protection Law (n 19) 112 ss.
831 Lynskey, The Foundations of EU Data Protection Law (n 19) 111. The question of the horizontal effect of the right to data protection has also been tackled by Hijmans, who notes that “an answer […] would clarify the nature of the right to data protection”: Hijmans (n 782) 61.
833 Lynskey, The Foundations of EU Data Protection Law (n 19) 118.
implementation in national law, application by national courts), and – while remaining a substantial dissimilarity – it does not seem to tell much about the substantial difference between privacy and data protection.

Additionally, as Hijmans points out, over the past decades, the ECJ (and then the CJEU) developed the “doctrine of direct effect”, according to which every EU law provision that is sufficiently clear, precise, and unconditional can be directly invoked before national courts. That may further challenge the possibility of distinguishing privacy and data protection’s rationale by virtue of their direct applicability to private relations. If one admits that the Charter and the Treaties may have direct effect between private parties then it is hard to see how insights about the difference between privacy and data protection, as embodied in Art. 7 and Art. 8 of the Charter and Art. 16 TFEU and Art. 39 TEU, could be sought in the difference in their horizontal scope. To be fair, it seems however debatable whether the EU primary law provisions in question would pass the justiciability test, i.e. the determination of whether they are sufficiently clear, precise, and unconditional to have direct effect between private parties. Plausibly, any sort of horizontal effect that the privacy and data protection provisions in the Charter and in the Treaties can have would derive from the primacy of EU law, which forbids interpretation of national law contrary to its spirit, rather than from their horizontal applicability.

All in all, the difference between the horizontal effects of the right to privacy ex Art. 8 ECHR vis-à-vis EU data protection does not however appear to be particularly meaningful in determining, by comparison, what constitutes data protection’s rationale. The same considerations seem to hold true, for the horizontal effect of Art. 7 of the Charter. As Kokott and Sobotta point out, ultimately, the differences in the personal scope of privacy and data protection result from data protection’s implementation in a secondary framework, rather than from the applicability of Art. 8 ECHR and Art. 7 and 8 of the Charter to horizontal relations.

4.2.2 Different Material Scopes

Considering the differences in the two rights’ material scope turns out to be more fruitful. Not all kinds of data processing seem to have the potential to breach the right to information privacy, while data protection has a wider and ‘agnostic’ scope. When it comes to the information falling within the scope of the right to privacy, the ECtHR recognises that all kinds

834 See CJEU, NV Algemene Transporten Expeditie Onderneming van Gend & Loos v Netherlands Inland Revenue Administration. Case 26-62, 5 February 1963, ECLI:EU:C:1963:1, and the supporting jurisprudence that followed.

835 On the potential for direct applicability of Art. 8 of the Charter and Art. 16 TFEU, see Hijmans (n 782) 61 ss.

836 As e.g. Hijmans seems to do: Hijmans (n 782) 61.


838 See Muir (n 837).


840 Contra, see Hijmans (n 782).

841 Juliane Kokott and Christoph Sobotta, ‘The Distinction between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR’ (2013) 3 International Data Privacy Law 222, 225.
of personal data can potentially interfere with the scope of Art. 8 ECHR, and yet never actually ruled that all kinds of personal data processing, regardless of its nature and contextual elements, fall under Art. 8 ECHR. The ECtHR, while recognising that even public data can interfere with the right to privacy, seems to consider, when assessing violations of Art. 8 ECHR, that the latter covers the processing of publicly available information only if the processing is systematic in nature, or reasonably unexpected by the data subject.

Data protection, on the contrary, applies to personal data – any information relating to an identified or identifiable natural person – regardless of whether private or public. To determine whether a person is identifiable, directly or indirectly, one must consider all the means reasonably likely to be used, either by the controller or by another person, to identify the data subject. To ascertain whether those means are “reasonably likely” to be used to identify the data subject, all objective factors (e.g. the costs and amount of time required, and the available technology) should be taken into consideration, unless practically impossible or illegal. It is immediately apparent how the notion of personal data in EU law, particularly in light of the very broad reading given by the CJEU’s jurisprudence, goes well beyond what is conventionally meant as ‘private’ information and what is traditionally associated with the concept of private life.

Furthermore, the meanings of ‘processing’ and ‘information’ adopted by EU data protection law are practically all-encompassing, so that any treatment of the data considered

---

842 The ECtHR, in underlining the extensive interpretation to be given to the concept of private life with respect to personal data processing under the ECHR, has been pointing to Convention 108: see e.g. ECtHR, Amann v. Switzerland. 30 EHRR 843 (2000) §65-67.

843 Hustinx (n 663) s2.D.

844 See e.g. ECtHR. Rotaru v. Romania (n 668) §43: “public information can fall within the scope of private life where it is systematically collected and stored”. The Court also acknowledges that permanent storage of otherwise mundane information can also bring the processing under the scope of Art. 8 ECHR: “The monitoring of the actions of an individual in a public place by the use of photographic equipment which does not record the visual data does not, as such, give rise to an interference with the individual’s private life […] On the other hand, the recording of the data and the systematic or permanent nature of the record may give rise to such considerations” – ECtHR, Perry v. United Kingdom, 39 EHRR 3 (2004) §40.

845 GDPR, Art. 4(1). On the expansive reading of the concept of personal data given by the Court, see CJEU, Patrick Breyer v Bundesrepublik Deutschland, Case C-582/14, 19 October 2016, ECLI:EU:C:2016:779 (on identifiability), and Peter Nowak v Data Protection Commissioner, Case C-434/16, 20 December 2017, ECLI:EU:C:2017:994 (on the meaning of ‘relating to’), which expands on YS and Others, Joined Cases C-141/12 and C-372/12, 17 July 2014, ECLI:EU:C:2014:2081. See also Bodil Lindqvist, Case C-101/01, 6 November 2003, ECLI:EU:C:2003:596.


847 GDPR, Recital 26.

848 Ibid.

849 See CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845).


851 GDPR, Art. 4(2): “any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means”. The ECtHR does however admit that mere storage of information is capable of interfering with the right to private life (see e.g. ECtHR, Leander v. Sweden, 9 EHRR 433 (1987) §48), which points toward a partial convergence between ‘processing’ in EU data protection law and the kinds of processing considered in the scope of Art. 8 ECHR.

integrates the legal definition of data processing under Art. 8 CFR, but might not be considered under the scope of Art. 8 ECHR (and hence 7 CFR). The Österreichischer Rundfunk case is an early example of the developing distinction between the respect of private life and the protection of personal data by virtue of their scope. The Court, after stating that the DPD must be interpreted in accordance with Art. 8 ECHR when applied to an area within its scope, divides the processing operations that can breach Art. 8 ECHR from the ones that cannot. A national law compelling a certain category of data subject to supply salary data to a governmental body was found to be a compression of Art. 8 ECHR, while the processing of the same data by the employer for its own internal purposes did not, provided that data protection rules were respected.

The right to private life, which has always been voluntarily left undefined by the ECHR, has become, over time, almost a general personality right, covering the broadest range of situations, only a subset of which involves information processing. However, the facet of the right to privacy ex Art. 8 ECHR that involves the processing of personal data (data privacy) has a narrower scope than the right to data protection, since it either requires the information processed to be somehow private in nature, or the processing to be systematic in kind. Privacy, in a nutshell, does not necessarily extend to all kinds of personal data processing, as data protection does. Either the processing of data or the data itself are “excluded from the privacy scope when (1) the data as such are not considered as private, (2) when there are no systematically stored [...] data, (3) when the data are not systematically stored with the focus on the data subject and (4) when the data subject could reasonably expect the processing”. Hijmans, while recognizing the narrower scope of data privacy under the ECHR, contends that, considering the right to privacy ex Art. 7 CFR, such conclusion does not hold true anymore, and that since the Charter and the Lisbon Treaty “all processing of personal data has a potentially adverse effect on the right to privacy under Article 7 Charter”. The author

---

854 Hustinx (n 663) s3.C.
855 CJEU, Österreichischer Rundfunk and others (n 853) §68.
856 CJEU, Österreichischer Rundfunk and others (n 853) §74.
857 See e.g. ECtHR, Pretty v. United Kingdom (n 300) §61; ECtHR, Niemietz v. Germany (n 300) §29; Peck v. United Kingdom (n 300) §57.
859 See Koops and others (n 263).
860 See e.g. CJEU, Österreichischer Rundfunk and others (n 853) §74, where the Court states that “the mere recording by an employer of data by name relating to the remuneration paid to his employees cannot as such constitute an interference with private life” (while also finding that “the communication of that data to third parties [...] infringes the right of the persons concerned to respect for private life, whatever the subsequent use [...] and constitutes an interference within the meaning of Article 8 of the Convention”).
861 Lynskey, The Foundations of EU Data Protection Law (n 19) 122 ss; De Hert and Gutwirth (n 236) 24; Kokott and Sobotta (n 841).
862 De Hert and Gutwirth (n 236) 26.
863 Hijmans (n 782) 42.
864 Hijmans (n 782) 47.
however also recognizes that “interference with the right to privacy is assessed in a contextual manner”, thus implicitly acknowledging how circumstantial factors about e.g. the way the data is processed, or the subjective elements of the data subject, might still exclude the processing instance from the scope of the right to privacy – but not from data protection’s one.

4.2.3 A DIFFERENT SET OF (INFORMATION) RIGHTS

The substantive protection granted by privacy and personal data protection, albeit partly overlapping, is another element that can allow to distinguish the two rights’ rationale: not all rights afforded by data protection law have been (or can be) mirrored in Art. 8 ECHR. Many rights granted by secondary data protection legislation have been incorporated in the ECHR’s privacy (rectius, private life) framework by the ECtHR’s jurisprudence, but the substantive safeguards granted by data protection law are still more varied and heterogenous. Some provisions of the DPD, such as the ones about automated decision-making or (to some extent) the security of the processing, were arguably already extraneous to the scope of Art. 8 ECHR (and of Art. 7 of the Charter), a postulate that has been reinforced by the introduction of additional data subject rights (e.g. personal data portability) by the GDPR. It is possible, largely due to the indeterminate nature of privacy, to make an argument for the inclusion of substantive protections like data portability and automated decision-making under the privacy umbrella, perhaps by linking them with the concept of informational self-determination. It however seems both easier and overall more consistent to just acknowledge how the rationale underlying many data subject rights has little to do with the respect for individuals’ private life, and is rather about the protection of a wider array of rights and freedoms that can be dented by personal data processing. The same holds true for the other side of the coin: the set of obligations EU data protection law mandates to data controllers. Their substantial content largely revolves around elements of accountability and procedural fairness (see e.g. Art. 30 of the GDPR) that may be made to fit under privacy’s ever-expanding umbrella, but that are much more likely to suit data protection’s instrumental and procedural nature, rather than private life’s substantive one.

The right to personal data protection substantiates itself in a wide array of diverse information rights, a hodgepodge whose coherence is given by the dual objective of safeguarding data subject’s rights and freedoms, and ensuring the free flow of data within the Union. The market-making framing of the DPD has been largely due to the ‘treaty-based games’ through which the EU managed to legislate on fundamental rights before having the competence to do so. However, the construction of EU data protection secondary law as (also) a market-making tool

---

865 Hijmans (n 782) 47.
866 See Lynskey, The Foundations of EU Data Protection Law (n 19) 127. Concrete examples are e.g. the right to data portability, or the right not to be subject to automated decision-making.
867 GDPR, Art. 22.
868 GDPR, Art. 32.
870 Janciute (n 787).
has arguably had an impact on the substantive safeguards offered by data protection law, on its differentiation from privacy, and ultimately on its standing as a different autonomous right.

4.2.4 DATA PROTECTION AS A PROCEDURAL TOOL OF TRANSPARENCY

A contribution by De Hert and Gutwirth identifies another fundamental difference between privacy and data protection, distinguishing them on teleological grounds by reference to their function within the democratic constitutional state. The authors define privacy as a ‘tool of opacity’, which protects individuals against interference by the state and by private actors by requiring abstention from undesired intervention. Privacy curtails power by setting normative limits to it. Data protection, on the other hand, is mainly a ‘tool of transparency’, regulating and channelling the exercise of power, rather than stopping it.

The distinction advanced between privacy as a tool of opacity and data protection as a tool of transparency is convincing and well grounded, and possibly one of the soundest justifications of the dichotomy between the two rights. It does not, however, entirely pinpoint the rationale of the right to data protection. As the authors themselves recognise, data protection is mainly a ‘tool of transparency’, but sometimes its substantive provisions curtail the possibility to process personal data, or set limits to the kinds of processing that can be done on personal data. As an example of the latter, the authors identify the rules applicable to sensitive data, profiling, and the purpose specification principle. Likewise, as the authors acknowledge, privacy has mainly a negative role, stopping the exercise of power, but also a positive one, ensuring individual freedom and self-determination. De Hert and Gutwirth’s theory is persuasive, and yet its main utility is to clarify the teleological dichotomy between privacy and data protection, rather than the latter’s substance.

The different nature of privacy and data protection could be highlighted, finally, by reference to substantive and procedural rights. Substantive rights are created to protect and uphold interests considered important, while procedural rights appear at a later stage, setting the conditions through which those substantive rights are implemented. Privacy, it is argued, is largely upheld through substantive provisions, while data protection has an eminently procedural nature. Data protection is meant to enable a broad range of rights and freedoms, privacy amongst others, through a varied range of provisions. It has been stated, in this respect, that it would only exist to serve other rights, and “does not directly represent any value or

871 See De Hert and Gutwirth (n 858).
872 De Hert and Gutwirth (n 858) s 3 and 4.
873 De Hert and Gutwirth (n 858) s 4.3.
874 De Hert and Gutwirth (n 858) s 3.2.
875 “The sheer wordings of the data protection principles [...] already suggest heavy reliance on notions of procedural justice rather than normative (or substantive) justice”: De Hert and Gutwirth (n 858) s 4.2.
877 “Article 8 of the Convention is no place for procedural questions. The framers of the Convention have designed other articles for that purpose. The transformation of Article 8 into a source of procedural rights and procedural conditions takes it away from the job it was designed for, viz. to prohibit unreasonable exercises of power and to create zones of opacity”: De Hert and Gutwirth (n 858) s 5.4.
878 See de Andrade (n 876); De Hert and Gutwirth (n 858); De Hert and Gutwirth (n 236); Claudia Quelle, ‘Privacy, Proceduralism and Self-Regulation in Data Protection Law’ [2017] Teoria Critica della Regolazione Sociale.
interest per se”. While the following section disputes that the right to personal data protection does not represent a value or interest on its own, its eminently procedural nature appears hardly disputable. Hustinx, in this respect, highlights that data protection is meant to provide a proactive, structural protection of the rights and freedoms that might be dented by personal data processing, regardless of a possible interference with the right to respect for private life.

The EU right to personal data protection is thus, formally and substantively, not privacy anymore. Its (quasi)constitutionalisation by the Charter formally separated from the right to privacy, and its evolution in the jurisprudence of the ECJ and the ECtHR and in EU secondary legislation made it essentially different, albeit still clearly connected to privacy. Data protection differs from privacy by virtue of the width of its scope and of the number of individual rights it grants. It is also distinct from privacy in that it is, mostly, a tool of transparency, rather than a tool of opacity. Data protection also has a prevalently formal and procedural nature, rather than a substantive one. Those traits may help delineate the essence and rationale of the right to data protection, but are not enough to do so on their own.

4.2.5 ON THE STRUCTURE OF THE CHARTER

The content and structure of Art. 8 CFR are also helpful in enucleating the constitutive elements of the right to data protection, and thus its rationale and, perhaps, its essence as well. In its initial paragraph, Art. 8 states, quite generically, that “(e)veryone has the right to the protection of personal data concerning him or her”. The first paragraph grants a right – the protection of personal data – but does not specify its content. The second paragraph details that “(s)uch data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law”, and that “(e)veryone has the right of access to data which has been collected concerning him or her, and the right to have it rectified”. The third and last paragraph of Art. 8 mandates that “(c)ompliance with these rules shall be subject to control by an independent authority”. The Charter’s Art. 8 lends itself to two possible readings: either a positive, permissive conception of data protection, or a negative, prohibitive one. The core of the distinction is the interpretation of the second and third paragraphs of Art. 8: do they detail the content of the right to data protection, or is the content of the right given by the first paragraph, while the second and third paragraphs detail the conditions under which such right can be limited?

The structure of the Charter would seem to point towards the former, framing data protection as a permissive right. The Charter lists the rights it contains in its first 50 articles, while Art. 51-

---

879 de Andrade (n 876) 97.
880 Hustinx (n 663) s.4.B.
883 In other words: is the processing of personal data an interference with the right to personal data protection in itself, to be legitimated through Art. 8(2) and 52 CFR, or does the processing of personal data become a compression of the right to data protection only when the conditions under Art. 8(2) are not met, and the conditions legitimising an interference are to be found only in Art. 52?
54 contain a number of general provisions governing the interpretation and application of the Charter. Art. 52, “Scope and interpretation of rights and principles”, provides for the conditions under which those rights can be limited. This construction echoes the binary nature of Art. 8 ECHR, in which the right is asserted in the first paragraph, and its admissible limitations in the second. It is also consistent with the parallelism between Art. 7 CFR, which enshrines the right to privacy, and Art. 8(1) ECHR: the right is contained in Art. 7 CFR and in Art. 8(1) ECHR, and the limitations are regulated by Art. 52 CFR and by Art. 8(2) ECHR, respectively.

Conceptualising data protection as a restrictive right, and thus Art. 8(2) as special conditions that legitimise an interference, rather than parts of the right itself, would arguably deviate from the binary structure of the Charter. That is also consistent with the fact that, as reported by Fuster and Gutwirth, early appearances of what would later become Art. 8 CFR were initially connected not to privacy, but to the right of access to documents. That would lean towards the conceptualisation of Article 8(2) CFR as part of the right to data protection, as it mentions access and rectification of personal data, rather than as (special) conditions justifying an interference, like Art. 52 does. The authors also report that all three paragraphs of Art. 8 CFR were initially envisioned as a single paragraph, an additional reason for framing Art. 8(2) and Art. 8(3) CFR as parts of the right itself, rather than as additional conditions legitimising an interference. That is also coherent with Convention 108’s approach, which “is not that processing of personal data should always be considered as an interference with the right to privacy, but rather that [...] processing of personal data must always observe certain legal conditions”. Furthermore, the Convention tasked with the Charter’s development reportedly considered making Art. 8 about the right to informational self-determination, thus espousing a prohibitive conception of personal data protection, but such approach was eventually rejected.

In accordance with a positive reading of the right to personal data protection as a tool of transparency, this chapter interprets Art. 8 CFR as being unitarily prescriptive of the content of the right, as opposed to framing Art. 8(2) and 8(3) as conditions legitimising an interference with the right. Consequently, it argues the components of Art. 8(2) and 8(3) should be seen as strongly linked with the essence of the right itself. The fairness principle, the purpose specification principle, the lawfulness principle, the right of access, the right to rectification,

---

884 “Everyone has the right to respect for his private and family life, his home and his correspondence” – Art. 8(1) ECHR.
885 “There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others” – Art. 8(2) ECHR.
886 Fuster and Gutwirth (n 882) 535.
887 Fuster and Gutwirth (n 882) 535.
888 Hustinx (n 663) s2.C.
889 Hustinx (n 663) 4.B.
890 The CJEU’s jurisprudence is however admittedly inconclusive in this respect, and sometimes seems to frame personal data processing as an interference with the right in itself, justifiable through Art. 8(2) and 52(1) CFR: see e.g. CJEU, Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others (n 767) §21, 36.
891 On the elusive content of the fairness principle see Damian Clifford and Jef Ausloos, ‘Data Protection and the Role of Fairness’ [2018] Yearbook of European Law.
and the control by an independent authority\textsuperscript{892} are thus constitutive elements of the EU right to personal data protection, rather than conditions, additional to the ones set by Art. 52 CFR, legitimising an interference with it. The figure below illustrates this position.

<table>
<thead>
<tr>
<th>Right &amp; source</th>
<th>Content of the right</th>
<th>Admissible limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Privacy &amp; Data Protection (ECHR)</td>
<td>8.1 ECHR</td>
<td>8.2 ECHR</td>
</tr>
<tr>
<td>Data Protection (Charter)</td>
<td>8.1 Charter</td>
<td>8.2 Charter</td>
</tr>
<tr>
<td>Privacy (Charter)</td>
<td>7 Charter</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1: Content and limitations to privacy and data protection in ECHR and Charter.**

It appears however debatable whether the lengthy and varied list of rights and principles contained in Art. 8(2) and 8(3) embodies data protection’s rationale, and (even more so) whether it directly constitutes data protection’s essence. If anything, Art. 8’s elements confirm the ties that the fundamental right to data protection has with the international instruments and the secondary legislation that preceded its (quasi)constitutionalisation, and its eminently procedural nature. The following section argues that the actual essence of the right to data protection is to be sought in a systematic, teleological interpretation of Art. 8 CFR, in light of (and as distinct from) Art. 8 of the ECHR and Art. 7 CFR, and with due consideration to the peculiar historical characteristics of the right’s development.

**4.3 A RIGHT TO A RULE**

The right to data protection has been strongly shaped by the evolution of the secondary legislation that has been enacted, since the early ‘70s, to tackle the ‘computers and privacy’ issue.\textsuperscript{893} Such legislation, in turn, has been entwined with, and responsive to, technological progress,\textsuperscript{894} so that its provisions have been changing in a substantial manner as time passed, and as new technologies impacted society. The increasing pervasiveness and societal importance that technology had since the turn of the past century led to the rapid emergence of a normative corpus, data protection law, which predated its constitutionalisation. In this sense, data protection legislation, and hence the fundamental right to data protection, responds to necessities that diverge, partly and yet sensibly, from the ones covered by the right to privacy, at least in the European milieu. This is reflected in the different scope data protection has vis-à-vis privacy. Narrower, as it only covers the processing of personal data, rather than all the whole spectrum of activities covered by privacy, but also broader, as it applies also to information not considered as private, that is not systematically stored or processed, or whose processing could reasonably be expected by the data subject. It is also

\textsuperscript{892} See Peter Hustinx, ‘The Role of Data Protection Authorities’ in Serge Gutwirth and others (eds), Reinventing Data Protection? (Springer 2009); Hijmans (n 782) 325 ss.
\textsuperscript{893} See Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19).
\textsuperscript{894} Hijmans (n 782) 48.
reflected in the narrower set of rights on information granted by privacy, in comparison to data protection.

The dichotomy between privacy as a tool of opacity, meant to curtail power, and data protection as a tool of transparency, meant to permit but channel through a set of rules, is also a key element in understanding the rationale of the right to data protection. It explains the reason of the constitutionalisation of data protection as a separate right, and is coherent with the functioning of secondary data protection legislation at a national and supra-national level. Data protection is meant to allow information sharing: there would be no need for it if there were a general prohibition of personal data disclosure, and the law very seldom prohibits the processing of personal data, rather mandating the requirements to be respected to make it lawful. This permissive conception of the right to personal data protection as a tool of transparency is also coherent with its procedural nature. Data protection channels a particular activity – personal data processing – by setting certain procedures and by granting a number of rights to data subjects and obligations to controllers and processors. Data protection aims at safeguarding a broad and diverse spectrum of rights and freedoms, privacy (inter alia) included.

If data protection were a negative, restrictive right, forbidding the processing of personal data, then every personal data processing instance would be a compression of such right, regardless of whether legitimate or not. In that case, the provisions ex Art. 8(2) and 8(3) CFR would be special legitimising conditions, identifying the requirements rendering a compression to the right to data protection legitimate. This chapter however adheres to a positive, permissive view of data protection, where personal data processing does not result in a breach of the right by itself. Accordingly, Art. 8(2) and 8(3) CFR and the elements it enumerates (the fairness principle, the purpose specification principle, consent and the other grounds substantiating the legitimacy principle, the right of access, the right to rectification, and the control by an independent authority) are seen as constitutive elements of the fundamental right to data protection, crucial elements to be considered in ascertaining its substance. That is not to say, however, that they constitute the essence of the right by themselves: it cannot be discounted that the CJEU might find other core elements of data protection, framing them as implied in Article 8(1) CFR. The list of rights and principles contained in Art. 8(2) and 8(3) CFR is too specific, reductive, and heterogenous to constitute the rationale of data protection on its own. Rather, it constitutes a synthetic representation of the societal posture that was developed, over the years and through several legislative instruments, as a response to the pervasiveness and importance that personal data processing acquired since the turn of the last century. To understand the essence and rationale of a right, one may also move from its infringement: a breach of the fairness, purpose specification, or lawfulness principle, of the rights to access

895 De Hert and Gutwirth (n 858); De Hert and Gutwirth (n 236).
896 Opinion of AG Alber in Case C-369/98 The Queen v Minister of Agriculture, Fisheries and Food, ex parte Trevor Robert Fisher and Penny Fisher, §41; Fuster and Gutwirth (n 882) 537. As Hustinx also stated, “The concept of ‘data protection’ was developed […] to provide legal protection to individuals against the inappropriate use of information technology for processing information relating to them. It was not designed to prevent the processing of such information or to limit the use of information technology per se. Instead, it was designed to provide safeguards whenever information technology would be used for processing information relating to individuals”: Hustinx (n 663) s1.
897 See de Andrade (n 876); Hijmans (n 782).
898 See Fuster and Gellert (n 882); Fuster and Gutwirth (n 882).
899 Hustinx (n 663) s4.B.
and rectification, of the independent authority’s control, does not necessarily amount to a breach of the essence of the right to data protection per se. There must be more to it than that: the CJEU often ruled on those matters without mentioning a breach of the essence of the right, nor a serious interference thereof.

It appears useful, at this point, to recapitulate what this chapter identified as constitutive elements of the right to personal data protection, as enshrined in the EU Charter of Fundamental Rights. To begin, data protection has had, since its very inception, a strong tie with the development of computing technologies. Such a tie led to the fast-paced development of a corpus of norms, that predated the (quasi)constitutionalisation of the right at EU level. EU data protection law emerged, in its distinction from privacy, as an omnibus regime, sector-neutral and applicable to public and private actors alike, whose broad scope encompasses an extremely wide range of processing instances. If privacy is prohibitive, a tool of opacity, data protection is permissive – a tool of transparency, designed to provide safeguards to individual whenever their personal data is processed. Data protection is also a procedural right, in that it hardly protects a specific interest on its own, but rather serves the goals pursued by other substantive fundamental rights. Finally, as data protection is a permissive right, the contents of Art. 8(2) and 8(3) CFR (the fairness, purpose specification, and lawfulness principles, the rights of access and rectification, and the control by an independent authority) are to be seen as constitutive of the right, rather than special conditions legitimising an interference with it.

This chapter’s conclusion is that the most coherent construction of data protection (as separate from privacy), at least in the European milieu, is simply as the fundamental right to having a (set of) rule(s) regulating the processing of personal data. While that may seem reductive, I believe it is not, and while that can prima facie appear generic, I argue that it is rather quite specific. Most importantly, I maintain that it is the most consistent explanation of the (quasi)constitutionalisation of the right to data protection as distinct from privacy, in light of its characteristics and of its sui generis development. In this respect, my argument echoes the notion of proceduralism in democratic theory, but applied, rather than to institutional legitimacy, to the identification of the value underlying personal data protection. The definition of ‘proceduralism’ has different nuances depending on the specific context of reference, but can be broadly defined as a theory that “justifies rules, decisions, or institutions by reference to a valid process, as opposed to their being morally correct according to a substantive account of justice or goodness”. Under this definition, proceduralism contrasts with epistemic theories of justice that ground the legitimacy of rules, decisions, or institutions by reference to

---

900 Charter, Art. 8(2).
901 Charter, Art. 8(3).
902 Lynskey, The Foundations of EU Data Protection Law (n 19) 15 ss. Lynskey identifies an additional characteristic of omnibus regimes: the supervision by an independent authority.
904 De Hert and Gutwirth (n 236); De Hert and Gutwirth (n 858).
905 de Andrade (n 876).
906 Fuster and Gellert (n 882); Fuster and Gutwirth (n 882).
908 Rocheleau (n 907) 906.
underlying values like knowledge and truth, however specified, rather than on the validity of the underlying procedure.\textsuperscript{909} Similarly, a ‘proceduralist’ view of data protection\textsuperscript{910} derives the latter’s legitimacy and status of fundamental right from the process through which it came into existence and from the rules through which it is substantiated, as both are expression of the collective will to regulate personal data processing. The value of data protection is, in a way, in the very existence of a system of rules and norms applying to the processing of personal data, regardless of its connection with concepts like privacy, or the secrecy and confidentiality of the information processed.

The European fundamental right to personal data protection emerged as a consequence of, and a response to, the rampant digitalisation of society, and the ever-growing importance of the processing of (personal) information. It embodies the collective\textsuperscript{911} decision of allowing personal data processing if and to the extent that it is subject to a certain set of rules, rather than prohibiting it unless specific requirements are satisfied. Such rules (and their judicial interpretation) have been substantially changing over time, evolving as the technological environment mutated. Their core is (partly) epitomised (yet not represented in its entirety) by Art. 8(2) and 8(3) CFR, whose content portrays a heterogeneous array of rights and principles that have remained staple regulatory mechanisms since the inception of EU data protection law. Data protection is “a catch-all term for a series of ideas with regard to the processing of personal data”,\textsuperscript{912} which makes the core of the right hard to pinpoint.

At the same time, such heterogeneity is indicative of the fact that data protection’s core goes beyond the elements enumerated by Art. 8(2) and 8(3) CFR. The Charter unequivocally granted the status of standalone fundamental right to data protection, explicitly distinguishing it from privacy, which, at least from an EU perspective, makes the position identifying data protection as a subset of the right to privacy hard to maintain. Considering data protection’s sui generis emergence\textsuperscript{913} and its (quasi)constitutionalisation by the Charter, this chapter contends that the most coherent reading of the fil rouge underlying the potpourri of disparate rights and principles that make up data protection is given by the societal will of regulating the processing of personal data by means of information technologies in itself. A ‘right to a rule’, so to speak,\textsuperscript{914}

\begin{itemize}
  \item \textsuperscript{909} However, the dichotomy between proceduralism and epistemic theories is not the only way to conceive proceduralism, and can be surpassed by coupling “a concern with democratic procedures with a concern with the epistemic quality of outcomes”: Fabienne Peter, ‘Pure Epistemic Proceduralism’ (2008) 5 Episteme 33. In other words, the idea that the legitimacy of a set of rules, decisions, and institutions can derive from the validity of the process from which they originate does not exclude an epistemic approach to the determination of what a valid process is. Such ‘epistemic proceduralism’ frames institutional legitimacy as deriving from the validity of the underlying democratic processes, and the validity of those processes through their epistemic value: their truth- and knowledge-seeking potential, however defined in the context of reference.
  \item \textsuperscript{910} The term ‘proceduralism’ has been previously used in connection with EU data protection law by Quelle, albeit with a different connotation – Quelle’s paper uses the term to describe the eminently procedural character of data protection: Quelle (n 878).
  \item \textsuperscript{911} This chapter uses ‘collective’, ‘societal’, and equivalent wording under the assumption that the mechanisms of (representative) democracy allow the expression of an underlying shared posture towards certain issues, in this case personal data processing.
  \item \textsuperscript{912} De Hert and Gutwirth (n 236) 3.
  \item \textsuperscript{913} I.e. its peculiar connection with technological development and the relation it has with the preceding secondary legislation.
  \item \textsuperscript{914} Hijmans, similarly, defines data protection as ‘the rules of the game’, a system of checks and balances based on a claim to fairness: see Hijmans (n 782) 59 ss. See also Christopher Docksey, ‘Articles 7 and 8 of the EU Charter: Two Distinct Fundamental Rights’ in Alain Grosjean (ed), \textit{Enjeux européens et mondiaux de la protection des données personnelles} (Éd Larcier 2015); Hustinx (n 663).
\end{itemize}
its rationale by now more akin to due process or effective judicial protection than it is to privacy.

That partly explains how the CJEU, in the Schrems case,\footnote{CJEU, Maximillian Schrems v Data Protection Commissioner. Case C-362/14, 6 October 2015, ECLI:EU:C:2015:650, §41.} after noting that “legislation permitting the public authorities to have access on a generalised basis to the content of electronic communications must be regarded as compromising the essence of the fundamental right to respect for private life”,\footnote{CJEU, Maximillian Schrems v Data Protection Commissioner (n 915) §94.} stated that “legislation not providing for any possibility for an individual to pursue legal remedies in order to have access [...] rectification or erasure of [personal] data, does not respect the essence of the fundamental right to effective judicial protection, as enshrined in Article 47 of the Charter”,\footnote{CJEU, Maximillian Schrems v Data Protection Commissioner (n 915) §95.} without however tying its reasoning with Art. 8 (or to its essence) directly. The Luxemburg court found attributes pertaining to data protection at the core of the right to an effective remedy\footnote{Maria Grazia Porcedda, ‘On Boundaries - Finding the Essence of the Right to the Protection of Personal Data’ in Ronald Leenes and others (eds), Data Protection and Privacy - The Internet of Bodies (Hart Publishing 2018) 283.} \textit{ex} Art. 47, so much that it considered the lack of the possibility to enforce them a breach of the very essence of the right. On one hand, the Schrems case highlights the ties that data protection has with neighbouring notions as procedural justice, democratic legitimacy, and the rule of law; on the other hand, it is regrettable that the CJEU did not tie Art. 47 and Art. 8 together, somehow leaving their connection to further interpretation. As it has been already noted,\footnote{Porcedda (n 918) 283.} both Art. 47 and Art. 8 CFR give expression to different facets of the rule of law, and it is unfortunate that the CJEU did not clarify their substantive relationship when it had the chance. Regardless, the reference to data protection provisions within the frame of Art. 47 CFR still highlights how data protection substantiates itself in a procedural system of heterogeneous checks and balances that is gradually detaching itself from privacy’s rationale.

4.4 \textbf{Theorising the Essence of the Right to Data Protection}

Defining data protection in the terms defined above – as the fundamental right to having a system of rules regulating personal data processing – leads to postulating, lacking authoritative interpretation, what the essence of the right to data protection may be.\footnote{Gellert and Gutwirth seem to contest the very fact that the rights to privacy and data protection have an essence (Gellert and Gutwirth (n 45)); that position does not however seem to be sustainable anymore, as the CJEU explicitly confirmed the existence of the essence of both rights in Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others (n 767) §39-40.} It also serves the purpose of distinguishing it from the right to privacy, respecting the dichotomy introduced by the Charter. In this regard, it has been noted how the CJEU has been (regrettably) assessing possible compressions of the right to data protection through the lenses of the right to privacy, following what has been defined as ‘privacy thinking’.\footnote{Fuster and Gellert (n 882); Hustinx (n 663) s4.D. Examples in case law are e.g. CJEU, Schecke and Eifert, Joined Cases C-92/09 and C-93/09, 9 November 2010, ECLI:EU:C:2010:662, and Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others (n 767) §39.} As plenty of authors already delved...
into the conflation between data protection and privacy by the CJEU, this section elaborates on this chapter’s view of the essence of the right to personal data protection.

The concept of essence, explicitly mentioned by the Charter, identifies one of the conditions to be fulfilled to justify an interference upon a right: any measure that interferes upon a fundamental right must respect that right’s essence. The concept of essence of rights is somehow pervasive within the EU multi-level fundamental rights framework: its source can be traced to many Member States’ constitutional traditions, to the CJEU’s pre-Charter jurisprudence, and to the ECtHR’s jurisprudence. There are three ways of looking at the concept of essence of a right. First, the essence can be seen as an inviolable core that cannot be legitimately interfered with: a boundary whose compression cannot be justified by overriding reasons. In this sense, the essence acts as an additional limit, and identifies the kind of interference with a right that cannot be legitimised through lato sensu proportionality. Second, the concept of essence can be seen, through the lenses of a relative stance, as the most serious kind of interference upon a right. Espousing this stance, the concept of essence identifies a kind of interference upon a right that can be legitimised through lato sensu proportionality reasoning, but only in cases where the trade-off between the rights and freedoms considered favours overwhelmingly the competing right. A third view of the concept of essence is of the limit past which an interference with the right leads to its extinction. Brkan depicts the possible interferences with fundamental rights as concentric circles, where one finds, at the outermost layer, no interference with the right, and then – progressing towards the centre – justified interferences, unjustified interferences, serious interferences, and interferences with the right’s essence at the very core. The distinction between interferences is particularly meaningful in light of the debate surrounding the relationship between balancing (proportionality sensu lato) and the essence of fundamental rights. A relative (or integrative) stance admits the possibility of compressing the essence of a fundamental right to safeguard another fundamental right; an absolute (or exclusive) stance

---

922 See e.g. Lynskey, ‘Deconstructing Data Protection: The “Added-Value” of a Right to Data Protection in the EU Legal Order’ (n 19); Fuster and Gutwirth (n 882); Fuster and Gellert (n 882); Tzanou (n 649).

923 Charter, Art. 52(1): “Any limitation on the exercise of the rights and freedoms recognised by this Charter must be provided for by law and respect the essence of those rights and freedoms” (italics added).


925 On the multi-level structure of the European fundamental rights’ framework, see Fabbrini (n 298).


929 Tridimas and Gentile (n 926) 803.

930 Maja Brkan, ‘The Concept of Essence of Fundamental Rights in the EU Legal Order: Peeling the Onion to Its Core’ [2018] European Constitutional Law Review 1, 332. In the Digital Rights Ireland case (CJEU, Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others (n 767) §39), for instance, the CJEU differentiates between a serious interference with the right to privacy and an interference with its core: “So far as concerns the essence of the fundamental right to privacy […] it must be held that, even though the retention of data required by Directive 2006/24 constitutes a particularly serious interference […] it is not such as to adversely affect the essence of those rights given that […] the directive does not permit the acquisition of knowledge of the content”.

931 See e.g. Robert Alexy, A Theory of Constitutional Rights (Oxford University Press 2009).

932 See e.g. Jürgen Habermas, Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy (2nd edn, MIT Press 1996).
constructs essence and proportionality as mutually exclusive concepts. Theorising the essence of the right to data protection is meaningful regardless of whether one adheres to the exclusive stance or the integrative one. In the former case, it identifies the interferences that are inadmissible regardless of the relative nature of the right to data protection; in the latter, it permits to nuance the gradation of the interference in view of the performance (ex ante) or the assessment (ex post) of the balancing test.

Brkan posits that the essence of a fundamental right is interfered with if 1) the interference threatens the very existence of that right, either for all right holders or for a specific right holder or group of right holders, and 2) if overriding reasons for such interference do not exist. The second part of the test espouses the absolute/exclusive position, which conceives interferences with the essence of the right as unjustifiable through proportionality balancing, and conversely – does not consider an interference with a right as a compression of its essence in all cases where there is possible to justify it by reference to overriding reasons (i.e. balancing). The first part of the test determines the presence of an interference with the essence of the right, which happens where the very existence of the right is called into question for all right holders (objective interference with the essence) or for a particular right holder (subjective interference with the essence). Brkan’s test to determine an interference with the essence of a fundamental right is particularly interesting in light of the conception of the right to data protection put forth by this chapter. The author postulates that the essence of a right is infringed where the existence of the right is called into question, and when there are no overriding reasons for the compression, as overriding reasons would introduce proportionality balancing, which is seen as incompatible with the concept of essence. Following such framework, and the framing of data protection outlined above, this chapter maintains that an infringement of the essence of the right to data protection is a compression strong enough to

933 Brkan, ‘The Concept of Essence of Fundamental Rights in the EU Legal Order: Peeling the Onion to Its Core’ (n 930) 359.
934 See the discussion in Mark Dawson, Orla Lynskey and Elise Muir, ‘What Is the Added Value of the Concept of the “Essence” of EU Fundamental Rights?’ (2019) 20 German Law Journal 763. While this chapter does not mean to participate to such a debate, it does implicitly adhere to an absolute approach, where an infringement of the essence of the right cannot be justified through proportionality balancing; a relative stance on the concept of essence is regarded as less compatible with the letter of letter of Article 52(1) of the Charter, with the spirit of EU member states’ constitutional traditions, and with the definition of some rights as fundamental: Tridimas and Gentile (n 926) 803.
936 Brkan, ‘The Concept of Essence of Fundamental Rights in the EU Legal Order: Peeling the Onion to Its Core’ (n 930) 363.
937 An objective interference with the essence results in the right being excessively restricted for all right holders, a subjective interference threatens the existence of the right for a particular right holder or group of right holders: Brkan, ‘The Concept of Essence of Fundamental Rights in the EU Legal Order: Peeling the Onion to Its Core’ (n 930) 350.
938 Brkan, ‘The Concept of Essence of Fundamental Rights in the EU Legal Order: Peeling the Onion to Its Core’ (n 930) 349 ss. Aside from objective and subjective interferences with the essence, the author also identifies absolute interferences with the essence when an absolute right is compressed. In a way, every compression of an absolute right is a compression of the essence of the right: “(i)n the case of fundamental rights that are absolute—such as the prohibition of inhuman or degrading treatment—the content of such rights is “all essence,” meaning that no limitation may be imposed on their exercise” - Lenaerts (n 924) 792.
threaten the very existence of the system of checks and balances in which the right to data protection is substantiated, when overriding reasons do not exist.\footnote{As Brkan points out, following an absolute/exclusive approach to the concept of essence of fundamental rights, the existence of overriding reasons through which the interference can be justified means that the interference does not compress the right’s essence, as proportionality balancing (i.e. overriding reasons) and interferences upon the essence of a right are mutually exclusive concepts: Brkan, ‘The Concept of Essence of Fundamental Rights in the EU Legal Order: Peeling the Onion to Its Core’ (n 930) 364. See also Maja Brkan, ‘The Essence of the Fundamental Rights to Privacy and Data Protection: Finding the Way Through the Maze of the CJEU’s Constitutional Reasoning’ (2019) 20 German Law Journal 864.}

Data protection is not an absolute right,\footnote{See Bonnici (n 935).} so it can be legitimately compressed by other overriding rights; it can also be illegitimately encroached upon, which would derive from a result of the balancing test favouring data protection rather than the other conflicting right. A violation of the essence of the right to data protection, which would be configurable (following an absolute/exclusive approach) only in the absence of overriding reasons justifying the compression, should be constructed only in cases where the functioning of the regime regulating personal data processing is called into question,\footnote{See CJEU, Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others (n 767) §40: “Nor is that retention of data such as to adversely affect the essence of the fundamental right to the protection of personal data enshrined in Article 8 of the Charter, because […] Directive 2006/24 provides […] that […] certain principles of data protection and data security must be respected”.} regardless of which specific provision is infringed, but rather having regard to the functioning of the system in its entirety.\footnote{See for instance CJEU, Opinion 1/15 of the Court (Grand Chamber). 26 July 2017, ECLI:EU:C:2017:592, §150, where the Court considers that “As for the essence of the right to the protection of personal data, enshrined in Article 8 of the Charter, the envisaged agreement limits […] the purposes for which PNR data may be processed and lays down […] rules intended to ensure, inter alia, the security, confidentiality and integrity of that data, and to protect it against unlawful access and processing”. The CJEU appears to hold that the presence of a set of rules (some of which it lists, while keeping the list non-exhaustive by inserting the ‘inter alia’ syntagm) is enough to avoid the configuration of a breach of the essence of Art. 8 CFR. The set of rules listed by the Court should not be read as a mere list of essential data protection principles, but rather as an indication that the compression of individuals’ right to data protection deriving from the PNR (Passenger Name Record) data sharing does not threaten the functioning of data protection’s system of checks and balances as a whole, as there still are “rules intended to ensure, inter alia, the security, confidentiality and integrity of that data”.} A violation of the essence of the right to data protection would thus exist in cases where the interference with the right challenges, explicitly or implicitly, the very societal choice of having an\textit{ omnibus} regime regulating the processing of personal data. Let us think of all the ways the system of checks and balances of which data protection is made can be practically broken by neglecting, exploiting, or misreading secondary law provisions that are not directly mentioned in Art. 8 CFR or defined as ‘essential’ by the CJEU, such as perhaps Art. 14(5) of the GDPR on the information to be provided where personal data have not been obtained from the data subject, or Art. 23 on the possibility to restrict the scope of (some of) the obligations and rights of the GDPR, or again the provisions about transfers of personal data to third countries or international organisations contained in Chapter V, or the provisions relating to specific processing situations in Chapter IX.

The rights and principles listed in Art. 8(2) and 8(3) CFR are amongst the earliest and most foundational instruments on which the EU’s data protection framework is based, and their violation\footnote{Particularly regarding the lawfulness principle.} would intuitions imply as a very plausible way to encroach upon the essence of
data protection. They do not constitute the essence on their own, but represent and embody the decision of regulating personal data processing at a fundamental rights level: it is such collective will that constitutes the essence of the right to data protection. A recent contribution by Porcedda\textsuperscript{944} follows a different route and attempts at enucleating the data protection’s essence by identifying a number of attributes – “the intrinsic and distinctive substantive dimensions of a right”\textsuperscript{945} – drawn from a number of sources,\textsuperscript{946} equating them with the ‘essential components’\textsuperscript{947} of the right.

Porcedda sees the essence as the core of an essential component (or attribute), submits that there may be multiple essences to data protection, and extracts those manifold essences\textsuperscript{948} from the relevant attributes. Enucleating the essential components of data protection is a useful exercise to identify potential serious interferences with the right, and the compression of one of the essential components of the right makes for a cogent case for a breach of the essence. However, the ‘essences’ identified by Porcedda seem to be mainly a further abstraction of data protection attributes/essential components, rather than its very essence. On one hand, the fact that data protection may have several different essences is intuitively unsatisfactory, as it gives up on finding a single, unitary rationale underlying data protection, a whole who should be more than the sum of its parts. Moreover, data protection is a qualified right, and can be compressed by competing rights: many of the ‘essential elements’ listed (e.g. the ones deriving from 8(2) CFR) can be legitimately compressed by competing rights, if the conditions set by Art. 52 CFR are met. If those attributes represented data protection’s essence(s) their compression would not be allowed, which does not appear to actually be the case. There is, finally, an additional possible level of abstraction that can be reached by tying together the ‘essential elements’ identified by Porcedda, the substantive characteristics of data protection, and its evolution in light of the function it gradually assumed in society. That last, essential abstraction level is given by the functioning, at a systemic level, of the checks and balances in which data protection is substantiated, as it represents the expression of a collective\textsuperscript{949} stance towards personal data processing.

The essence of the system of checks and balances in which data protection is substantiated – our fundamental ‘right to a rule’ regulating personal data processing – should be seen as the collective decision of generally allowing the processing personal data by virtue of its promises, while at the same time regulating it on account of its perils. An interference with the essence of the right to data protection is thus different from a ‘regular’ interference, regardless of how ‘serious’. While the former dents part of data protection’s system of checks and balances, the latter questions and endangers the very functioning and legitimacy of the collective posture

\footnotesize{\textsuperscript{944} Porcedda (n 918).}

\footnotesize{\textsuperscript{945} Porcedda (n 918) 289. The attributes identified by Porcedda are legitimate processing (1), oversight (2), supervisory authority (3), human intervention (4), data subjects’ rights (5), and security and minimisation (6) - see Porcedda (n 918) 310.}

\footnotesize{\textsuperscript{946} I.e. EU primary and secondary law, post-Charter CJEU jurisprudence, ECtHR jurisprudence in light of Convention 108, doctrine, and the Fair Information Processing Principles (FIPPs).}

\footnotesize{\textsuperscript{947} Wording used in CJEU, Maximilian Schrems v Data Protection Commissioner (n 915) §41.}

\footnotesize{\textsuperscript{948} Porcedda (n 918) 310. The essences identified in the CJEU’s jurisprudence and the ones proposed by Porcedda are purpose limitation, ex-post supervision, human intervention, a mixture of data subjects’ rights (access, rectification, right to object) and the provision of security safeguards.}

\footnotesize{\textsuperscript{949} Again, framing data protection as a collective and shared social stance moves from the assumption of the legitimacy of the underlying democratic decision-making processes.}
towards personal data processing in its entirety, and hence ultimately its deepest roots: the rule of law, and the democratic legitimacy of EU regulation.

4.5 WHAT DOES THIS MEAN FOR OPEN DATA?

The previous chapter mentioned how many stakeholders within the open data ecosystem perceive privacy legislation as a barrier to open data sharing and re-use. It also highlighted how, more often than not, the relevant regulatory barrier is not the right to privacy, but rather the one to data protection. The preceding sections followed up by accounting for the difference between the right to privacy and the one to data protection, and for the latter’s emergence and development as a standalone fundamental right in the EU.

“Same difference”, one might however argue – in the end, there is a secondary legislative framework that functions as a barrier to open data sharing and re-use, regardless of whether the underlying fundamental right is privacy or personal data protection. That is hard to dispute – data protection law does curtail the possibility to share and re-use personal data as open data. However, as the previous chapter explained, open data is not a right nor a value on its own; it is rather an extremely permissive data sharing strategy that aims at fostering, as much as possible, information publishing and re-use, and that is backed up at a policy and at a legislative level. If one considers the whole spectrum of information sharing and usage options, rather than just open data – the most extreme end of the openness spectrum, so to speak – then the distinction between privacy and data protection becomes paramount to understand why, rather than just whether, data protection curtails open data.

Indeed, as remarked above, data protection is an inherently permissive regime, meant to channel personal data processing through a set of rules rather than to prohibit it tout court. It is, so to speak, a tool of transparency, as opposed to privacy, which shields from the exercise of power, and can thus be described as a tool of opacity. It is about the protection of personal data from unbridled processing, rather than about the respect for individuals’ private life. Data protection law regulates the processing of a subset of all information – personal data – not by prohibiting it, but by setting rules and principles about how to do it. In this sense, personal data protection and data processing are not inherently at odds; the picture does however change if, rather than data sharing and use in general, one considers open data in particular.

Even when framing data protection as a permissive, procedural, and auxiliary right – a tool of transparency that defines how personal data is processed, rather than whether – the system of checks and balances in which such right substantiate clash with the radical openness underlying the concept of open data. As discussed in the previous chapter, actual open data (as opposed to less permissive data sharing and re-use policies) is fundamentally incompatible with mechanisms such as access controls, use restrictions, logging and record-keeping, which are central instruments in data protection. Likewise, the array of data subject rights and the

951 De Hert and Gutwirth (n 858); De Hert and Gutwirth (n 236).
952 Cfr. the wording of Art. 8 ECHR and Artt. 7 and 8 of the Charter.
953 See Pollock (n 635).
954 See GDPR, Chapter III.
corresponding data controller (and processor) obligations are hard, if not practically impossible, to exercise and comply with when the data whose processing they were made to channel is shared as widely, and re-used as freely, as open data is meant to be. Data protection is a system of checks and balances regulating personal data processing. The utter absence of constraints required by the concept of open data – digital information that is public, interoperable, machine-readable, and freely accessible and re-usable by anybody and for any purpose – is inherently at odds with those checks and balances.

For the most part, the emergence of data protection legislation answered to the same societal push that lead to the development of the open data movement: the rise of the information and communication society, and the exponential pace of the development of computing technologies. While personal data protection materialised on account of the ICT society’s perils, open data arose by virtue of its promises – serendipitous discovery of new knowledge, and value generation deriving from the access to and re-use of information that was traditionally kept closed by virtue of rights, interests, and legal regimes other than privacy and data protection. The rationale underlying the rise of the open data movement has to do with intellectual property rights, business confidentiality, and organisational cultures that favour ownership and secrecy over sharing and openness, rather than with individuals’ right to privacy, the respect for private life, and the existence of a system of rules regulating the processing of personal data. As the previous chapter highlighted, the laws and policies supporting open data have always been meant not to be prejudicial to privacy and data protection: in a nutshell, open data and data protection run on parallel tracks, and their regulatory convergence (and possible conflicts) are solved by giving preference to the latter.

That does not however mean that personal data cannot be shared and used at all, but just that they cannot be published and re-used as actual open data. The rights and principles on which data protection is based are practically incompatible with the uncompromising lack of limitations on which open data is based, but the very same information can often still be shared and used, in its ‘raw’ form or after it gets processed into ‘safer’ data, in compliance with the applicable regulatory regimes. In a way, the idea of balancing open data and data protection is tantamount to the idea of deregulating personal data processing: data protection does not impede personal information processing, it just sets out rules on how to do it. In other words, the idea of balancing open data and data protection (as opposed to balancing data protection and the rights and interests underlying open data) is not about the alternative between keeping personal data ‘closed’ and rendering it accessible and usable, but rather about

955 See GDPR, Chapter IV.
956 See Open Knowledge International (n 32).
957 On the topic see Ceruzzi, A History of Modern Computing (n 99); Ceruzzi, Computing: A Concise History (n 99); Rid (n 678); Yasha Levine, Surveillance Valley: The Secret Military History of the Internet (PublicAffairs 2018).
958 Intellectual property, trade secrets and business confidentiality, database rights, market competition, public and national security are amongst the main normative reasons why data has been traditionally kept locked; limitations to the networking and data processing capabilities available amongst the technical ones. Organisational issues, management, and culture played (and still play) their part too.
959 The examples are manifold: the purpose specification principle, for instance, is fundamentally incompatible with the ‘usable for any purpose’ tenet of open data; the storage limitation principle is another example: albeit not necessarily incompatible at a conceptual level, it is impossible to achieve in light of the decentralised and distributed nature inherent to open data in practice.
960 E.g. through generalisation, aggregation, pseudonymisation, anonymisation, or other sanitization measures.
weakening or removing the checks and balances regulating how and to what extent personal data is accessible and usable.

Data protection is, indeed, not an absolute right — as every other qualified right, it must give way whenever it is outweighed by a countervailing right (e.g. freedom of expression, good administration) and its compression is deemed legitimate by the result of the balancing test performed, either ex ante or ex post. That is to say that whenever the rights and freedoms underlying open data (and not open data in itself, as open data is not a right on its own) happen to clash with the right to personal data protection, releasing the information in question as actual open data would never be a proportionate outcome for any balancing test. Even in cases when it clearly stands to reason that one’s right to data protection ought to be somehow compressed by other rights and freedoms, a proportionate outcome would still imply some conditions that restrict, at least to some extent, the possibility to access or use the information concerned; those conditions are inherently incompatible with the freedoms on which open data is based. The next chapter expands on the reasons why the idea of balancing open data and data protection is intrinsically contrary to the notion of balancing itself, both in its normative sense and considering the common meaning of the word. The chapter following it clarifies how suitable mechanisms to balance individuals’ right to personal data protection and the rights and freedoms underlying open data – but not open data per se – are already foreseen in data protection secondary legislation.

4.6 CONCLUSION

This chapter framed the fundamental right to data protection as the expression of a societal choice towards personal data processing. A permissive, procedural ‘right to a rule’, allowing – and yet channelling – an activity that is as fundamental for modern society as it is possibly dangerous. It tied the sui generis emergence of EU data protection to technological development, and its (quasi)constitutionalisation to the growing importance of the secondary legislative framework through which it has been enacted. In light of the heterogeneousness of the rights and principles substantiating the fundamental right both at primary and at secondary level, of its formal differentiation from privacy, and of its procedural and permissive nature, this chapter argued that the most coherent conceptualisation of the right to data protection is of a system of rules and principles regulating personal data processing by virtue of its potential impacts on individuals and society. The essence of the right to data protection has been framed, accordingly, as the collective, societal choice of having system of checks and balances regulating personal data processing. The infringement of the essence of the right to data protection has been defined as a compression strong enough to threaten the very existence of such a system of checks and balances when overriding reasons do not exist, regardless of which specific component of the right is compressed.

961 See Bonnici (n 935).
962 There are several examples of balancing tests in the European multi-level fundamental rights framework: from the ones enshrined in the EU member states’ constitutions, to the ECHR’s ones (e.g. Art. 8, Art. 11), to the Charter’s Art. 52. All of them share similar traits — at their core, they assess the necessity, legitimacy, and proportionality of the compression of a right.
963 See Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19).
964 See Docksey (n 914); Hustinx (n 663).
965 See de Andrade (n 876); Quelle (n 878).
966 See De Hert and Gutwirth (n 858); De Hert and Gutwirth (n 236); Fuster and Gutwirth (n 882).
EU data protection is an inhomogeneous assortment of rules and norms whose substance hardly fits under a unitary rationale. Technological development, the diffusion of computing, and the rampant datafication of society led to the development of a *sui generis* right that spun out of privacy into something different, new, and still in flux, which was then elevated to the status of fundamental right by the Charter. Data protection is the answer to the power and information asymmetries running between who controls the means of data processing and the individuals those data refer to, and responds to a recent need for protection that has risen in parallel with the advances in computing technologies and with their role in contemporary society.

In a way, the justification of the ascension of data protection to the status of fundamental right should not be sought in the (scarce) conceptual autonomy or systematic coherence of the heterogeneous array of rights and principles of which data protection is made. Rather, it is the (quasi)constitutionalisation of data protection as an autonomous fundamental right that helps delineating its substance and its difference from privacy. The choice of creating an *omnibus* system of checks and balances regulating any kind of personal data processing, and then to enshrine such a system at a primary level, was far from an obvious path. The seemingly incoherent set of rights and principles of which data protection is made, in light of its explicit (quasi)constitutionalisation by the Charter and moving from the assumption of the democratic legitimacy of the underlying legislative process, create a pragmatic system of protection that should be seen as the embodiment of such choice.

A consistent conceptualisation of the right to data protection serves the purpose of justifying the formal dichotomy between privacy and data protection introduced by the Charter, which are still too often conflated. It also boosts the legitimacy of the EU data protection framework by explicitly outlining how it expresses a definite societal posture towards personal data processing. Most importantly, recognizing the conceptual autonomy of personal data protection and its ties to the advance of computing technology and (hence) modern society could foster the development of Art. 8 CFR to its full potential. The law, both at a secondary and at a primary level, is an inherently living instrument, responsive to society’s changing needs. The Charter, through the introduction of Art. 8, created the possibility to further advance data protection into something able to steer society towards a place where the power and information asymmetry inherent to personal data processing is curbed by default. It would be a pity to let that chance go by keeping data protection under privacy’s shadow, and by focusing on its past emergence rather than on its capability to shape future society.

Regardless, even espousing a conception of the right to data protection as open and as permissive as the one adopted by this thesis – or perhaps particularly in light of how this thesis sees personal data protection – the idea of balancing it with open data remains difficult to justify. Data protection just sets rules on how personal data must be processed, and has always been meant to be balanced with the rights and freedoms underlying open data. Open data *per se* is a different issue: amongst all the possible data sharing and re-use options, it is the most uncompromising one; within all the possible shades in the openness spectrum,967 it is the most radical. The checks and balances in which data protection substantiates, despite not being meant to stop information processing, are fundamentally incompatible with the requirements defining open data968 (as distinct from other data sharing and re-use policies). The following chapter clarifies how the idea of balancing open data and data protection is, at its core,

967 See Open Data Institute (n 638).
968 See Open Knowledge International (n 32).
fundamentally incompatible with the very notion of balancing. It argues that, whenever there is the need to perform a balancing test between the right to data protection and the rights and freedoms that can be supported through open data sharing, open data itself cannot be a proportionate outcome. Proportionality *lato sensu* – balancing – would inevitably lead to compromise, to the choice of less permissive data sharing and re-use option than open data.
5. Balancing Data Protection and Open Data?

EU data protection law makes it impossible to release personal data as open data. The principles, rights, and obligations on which it is based are practically incompatible with the complete, utter openness required for information to be open data. At the same time, the scope of EU data protection law is very wide and, some argue, ever-expanding. It has hence been claimed that data protection law constitutes a major threat for the future of open data, and that open data and data protection thus need to be balanced. If more and more kinds of data become personal, less and less data can be released as open data, with purported detrimental effects on transparency, public-sector efficiency, and economic growth – the oft touted benefits deriving from open data sharing and re-use.

This chapter maintains that such argument – the idea that it is necessary to somehow ‘balance’ open data and data protection – is moot on different levels. First, it does not take into account the hierarchical relationship between data protection and open data at a (quasi)constitutional level. Second, it misconstrues the actual functioning of EU data protection law. Third, the idea of balancing open data and data protection does not consider how open data is but one of the many ways in which information can be shared – the most uncompromising way of releasing data. Fourth, it neglects the fact that mechanisms made to balance data protection and other rights and freedoms, including the ones underlying open data sharing and re-use, have (already) been embedded in data protection law from its earliest stages. Finally, and perhaps most importantly, the idea that it may be necessary to ‘balance’ open data and data protection moves from a non-technical reading of what balancing is from a normative perspective, or from a definition of ‘balancing’ that is extraneous to the European legal tradition.

---

969 “Any information relating to an identified or identifiable natural person”: GDPR, Art. 4(1).
970 On the importance of definitional clarity in open data, see Pollock (n 635).
971 I.e. data that “anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)”: Open Knowledge International (n 32).
972 GDPR, Art. 5.
973 GDPR, Art. 12-22.
974 Inter alia GDPR, Art. 24-39.
975 For instance, intellectual property rights over a dataset can be waived so that anyone can use it for any purpose, and yet if that dataset contains personal data then data protection’s purpose limitation principle applies, and the dataset would not then be reusable “for any purpose”.
978 See e.g. the SPOW project (STW-Maps4Society project number 13718). See also Zuiderveen Borgesius, Van Eechoud and Gray (n 41).
979 See e.g. Scassa (n 614); Zuiderveen Borgesius, Van Eechoud and Gray (n 41).
980 I borrow the term “(quasi)constitutional” from Gellert and Gutwirth (n 19).
I contend that, ultimately, arguing for ‘balancing’ between open data and data protection is not about balance at all, but rather about deregulation – about lowering the level of protection afforded to individuals in favour of data re-users. I conclude that the balance between open data and data protection is a non-existing problem, truly a red herring, and should not be tackled as such.

The argument is developed as follows: the next sections will continue by discussing the process of subsuming of a case under a (set of) norms – along with balancing, one of the two main operations of legal argumentation. Such clarification is a necessary precondition for a serious discussion about the idea of balancing open data and data protection, as – in the current EU legal framework – their interaction is not dealt with by balancing them, but rather by simply applying a set of relevant norms: by subsuming a case under a rule. In other words, de iure condito, proportionality reasoning has no place in regulating the interaction between open data and personal data protection; musing about the idea of ‘balancing’ open data and data protection has necessarily to be done de iure condendo. The chapter then resumes by clarifying what balancing and proportionality are in legal scholarship; while the two terms are often used interchangeably and to refer to their common meaning, balancing and proportionality have a specific significance in the context of legal reasoning. The chapter continues by discussing the extent to which the right of data protection would have to be compressed to allow the publication and re-use of personal as (actual) open data. It then considers whether the idea of balancing open data and data protection for the sake of the development of the smart city environment is (lato sensu) proportionate, i.e. whether it would be suitable to the goals pursued, necessary in a democratic society, and whether the benefits it would bring would be (stricto sensu) proportionate to its detrments.

5.1 On rules and principles

The role of proportionality within modern European law, jurisprudence, and doctrine is extremely prominent. Proportionality provides for an ideal conceptual framework when reasoning de iure condendo – as if positive law applicable to a specific case, and rules solving potential conflict of laws, did not exist. It is also the prime method of adjudication when dealing with norms that function as principles, i.e. optimization requirements meant to ensure the best possible trade-off between equally important competing rights. Lest readers think adherence to the law admits of degrees of noncompliance, this section explains the role of proportionality balancing by clarifying the distinction between norms that are principles and norms that are rules.

Norms have prescriptive character, as they dictate what ought to be. In this sense, both rules and principles are norms, as they both express commands, prohibitions, and permissions, and are backed up by the binding force of law. The distinction between norms that are rules and norms that are principles has been argued in legal scholarship for the better part of the past century. Here, I take the view advanced by Alexy, who frames principles as optimization requirements – norms which entail that something is to be realised to the greatest extent possible given the legal and factual possibilities. Rules, on the other hand, are norms that are

982 Alexy (n 931) 47.
either fulfilled or not: “fixed points in the factually and legally possible”. Every norm is either a rule or a principle.

The distinction between rules and principles matters particularly when norms clash: different principles might be competing, and diverse rules might be conflicting, giving rise to antinomies. The conflict between open data and data protection, to some extent, could be seen as a potential clash of norms. In those cases, the conflicting rules are mutually incompatible – their parallel application is impossible – as much as the pursuit of two conflicting principles at the same time would lead to inconsistent results. The solution of conflicts between norms is different when it comes to conflicts between norms that are rules, on the one hand, and of conflicts between norms that are principles, on the other hand.

A conflict of rules – an antinomy – is solved through one of two ways: by reading an exception into one of the rules in conflict, or by expunging one of the rules from the legal system of reference. The law is the law, and when two rules collide, one of them has to give way. Lacking explicit textual exceptions, several criteria – other rules such as e.g. hierarchy, the chronological criterion, speciality over generality – may be used to determine which rule is to be applied and which rule is to be neglected. A rule is either valid or invalid: tertium non datur. The adjudication of competing principles, on the other hand, does not result in one of the principles being invalid: it results in one of them being outweighed by the other, in light of the circumstances surrounding the concrete case. While conflicts of rules are solved by reference to their validity, the competition of principles is adjudicated by reference to their respective weight in the case at hand.

In other words, an antinomy of rules is solved through subsumption: the inclusion of a specific case under a general rule. Rules may conflict where a concrete case can, prima facie, fall under (i.e. be subsumed by) two conflicting rules, and no suitable exception can be read in one of them. In that case, the solution is given by a further subsumption (a meta-subsumption, if you will) under another set of norms, general criteria that can be used to prioritize a rule or another. Principles behave in quite a different fashion: after an initial subsumption, in which two competing principles are deemed both applicable to a certain situation (i.e. when a certain situation is subsumed by two competing principles), the principle to be prioritized is decided through balancing. The issue is not one of validity, as it happens when rules conflict, but one of optimization, as principles are optimization requirements that aim at ensuring the best possible trade-off within the realm of the legally and the factually possible.

I do not mean to delve into legal theory, constitutionalism, or the formalisation of judicial reasoning and argumentation, and yet this thesis would be remiss if it did not clarify that, right now and within the European legal framework, conflicts between open data and data protection would be resolved through subsumption, rather than through proportionality. The conflict between open data and data protection is not to be solved as a conflict of principles, a clash of optimisation requirements whose solution is reached by balancing them. It is to be solved by complying with a set of rules: the positive law regulating the matter is a clear and a well-established part of the EU legal system. In other words, the interaction between open

983 Alexy (n 931) 48.
984 “Lex superior derogat inferiori”.
985 “Lex posterior derogat priori”.
986 “Lex specialis derogat generali”.
987 See generally Alexy (n 931).
data and data protection does not need to be ironed out through a proportionality assessment, but is rather dealt with through the subsumption of a case under a precise set of rules.

5.2 SUBSUMING THE INTERACTION BETWEEN OPEN DATA AND DATA PROTECTION

The interaction between PSI laws and data protection is, in principle, an established part of EU law. The publication of PSI must comply with EU data protection law to its full extent, and PSI law is meant not to be prejudicial to data protection legislation. There have been, since the 2013 amendments to the PSI Directive, two provisions that make its relationship with data protection law explicit. The first one is an exception to the general PSI regime: documents access to which is excluded or restricted on the grounds of protection of personal data, and parts of documents accessible by virtue of those regimes which contain personal data whose re-use is incompatible with data protection law, are excluded from the scope of the PSI Directive. The second provision is a non-prejudice clause and clarifies, at a higher level, how the PSI Directive and EU data protection law interact, by specifying that the PSI Directive does not affect the level of protection of individuals with regard to the processing of personal data under EU and national law, and in particular does not alter the obligations and rights set out in the Data Protection Directive (now repealed by the GDPR).

Notably, the Commission’s Proposal for a recast of the PSI Directive, while maintaining the exception under Art. 1(2)(h), did not initially contain a provision equivalent to Art. 1(4), for which the PSI Directive is meant not to be prejudicial to data protection legislation. On one hand, the relationship between PSI legislation and personal data protection is well established in Union law, and thus the specification ex Art. 1(4) of the PSI Directive could have seemed unnecessarily redundant. However, an explicit clarification of the unfettered applicability of EU data protection law enhances legal certainty and consistency. Luckily, the EU legislature, heeding the European Data Protection Supervisor’s advice, ended up maintaining the non-prejudice clause under Art. 1(4).

Maintaining the non-prejudice vis-à-vis personal data protection has been a good course of action, as the obligations deriving from the application of EU data protection law are hardly compatible with the release of (personal) data as PSI. The right to personal data protection is a tool of transparency, meant to channel data processing through a set of rules – as opposed to the right to privacy, which is rather a tool of opacity, meant to curtail the exercise of power.

---

988 PSI Directive (recast), Art. 1(2)(h); formerly Art. 1(2)(cc) of the consolidated 2013 PSI Directive.
989 On the implementation of the PSI Directive’s exception and non-prejudice clause on grounds of personal data protection in Member States’ national legislation, see Andreas Wiebe and Nils Dietrich, Open Data Protection-Study on Legal Barriers to Open Data Sharing-Data Protection and PSI (Universitätsverlag Göttingen 2017).
990 Such omission was particularly noticeable as the non-prejudice clause ex Art. 1(4) has been in the PSI Directive since its first draft, while the exception currently under Art. 1(2)(h) was introduced with the 2013 amendments.
data protection is thus not meant to curb information processing per se. However, several of its substantive principles (from the lawfulness principle to the purpose specification and data minimization ones) and of the provisions granting specific rights to data subjects (from rectification to erasure) cannot be practically complied with unless the data controller (a public sector body or a private undertaking) engineers the processing by implementing constraining measures, such as access controls or use restrictions, which conflict with the rationale underlying PSI re-use.

That is to say that even if EU data protection law is (mostly) a permissive regime meant to allow personal data processing, and thus public sector bodies can process and publish personal data lawfully by structuring the processing correctly, the system of checks and balances it implements still constrains (personal) data processing in a way that is inherently at odds with the rationale underlying the PSI Directive – fostering (non-personal) data sharing and re-use. Excluding personal data from the scope of the PSI Directive in an explicit manner avoids unnecessary troubles in reconciling two different regulatory frameworks that are not meant to intersect.

However, the wording of the exception is fairly confusing in its reference to documents “access to which is excluded or restricted by virtue of the access regimes on the grounds of protection of personal data, and parts of documents accessible by virtue of those regimes which contain personal data the re-use of which has been defined by law as being incompatible with the law concerning the protection of individuals with regard to the processing of personal data”. The EDPS, in its Opinion on the recast of the PSI Directive, suggested an alignment with the wording of Art. 4(1)(b) of Regulation 1049/2001 regarding public access to European Parliament, Council and Commission documents, under which EU institutions can deny access to a document where it “would undermine the protection of privacy and integrity of the individual, in particular in accordance with Community legislation regarding the protection of personal data”. The EDPS also suggested to clarify the reference to “documents” and to “parts of documents”, or to just “refer to ‘regimes which contain personal data the re-use of which’ could undermine the fundamental right to privacy and the protection of personal data.”

Regrettably, the final wording of Art. 1(2)(h) of the 2019 PSI Directive maintained continuity with Art. 1(2)(cc) of the 2013 PSI Directive, and thus its convoluted wording referring to ‘documents’ and to ‘parts of documents’ and to their access regimes. The amendments that resulted in the new Art. 1(2)(h) merely introduced an additional subparagraph, which includes the sentence “[…] or as undermining the protection of privacy and the integrity of the individual, in particular in accordance with Union or national law regarding the protection of personal data”.

---

993 See De Hert and Gutwirth (n 858). On the difference between the fundamental right to privacy and the one to personal data protection, see e.g. Kokott and Sobotta (n 841); Tzanou (n 649). On the process of (quasi)constitutionalisation of the right to data protection and on its emergence within the EU fundamental rights system, see Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19); De Hert and Gutwirth (n 236).
994 See Fuster and Gutwirth (n 882).
996 European Data Protection Supervisor (n 991).
998 European Data Protection Supervisor (n 991) 7.
Similarly, the Access to Environmental Information (AEI) Directive provides that “Member States may provide for a request for environmental information to be refused if disclosure of the information would adversely affect [...] the confidentiality of personal data and/or files relating to a natural person where that person has not consented to the disclosure of the information to the public, where such confidentiality is provided for by national or [Union] law”. Once more, EU law allows for an exemption to the general AEI regime on data protection grounds, yet again with a confusing wording. The protection of personal data is not (just) about confidentiality, and the reference to consent – only one of the grounds available for the processing of personal data – can hardly be justified in light of the functioning of the EU data protection regime. Furthermore, the adverb “adversely” would intuitively lead to believe that, if the processing does not affect individuals in a negative manner, the exception to the AEI regime on data protection grounds cannot be claimed, which is surely not the case. Finally, Art. 4(2)(f) of the AEI Directive mentions “personal data and/or files relating to a natural person”: the latter is much broader than the notion of personal data, where the fact that the information must relate to a natural person is just one of the cumulative requirements of the notion. In any case, the meaning of the provision is pretty clear – the protection of personal data allows to raise an exception to the standard access regime.

The same level of protection is arguably foreseen at an infrastructural level – open data is not just about the information, but also about the environment and infrastructure around it. For instance, Recital 24 of the INSPIRE Directive, which is another brick of the EU open data ecosystem, explicitly states that “(t)he provision of network services should be carried out in full compliance with the principles relating to the protection of personal data in accordance with Directive 95/46/EC”. INSPIRE’s Art. 11(1)(f) follows suit, with a questionable formulation that echoes Art. 4(2)(f) of the AEI Directive, by allowing EU Member States to limit public access to spatial data and services where it would adversely affect “the confidentiality of personal data and/or files relating to a natural person where that person has not consented to the disclosure of the information to the public, where such confidentiality is provided for by national or [Union] law”. What can be said about the impact that open data is meant to have on data protection law can also be said, mutatis mutandis, about the infrastructure around it.

A similar regulatory approach can be found in other legislative instruments promoting, directly or incidentally, the development of open data in the EU. For instance, Recital 9 of the Copernicus Regulation underlines how the Copernicus satellite data programme “should be implemented consistently with other relevant Union instruments and actions, in particular [...] in the field of security, protection of personal data”. The ITS (Intelligent Transport Systems) Directive deals with data protection and the re-use of ITS-related PSI in the same Article. Before clarifying that the PSI Directive applies to the ITS domain, Art. 10 of the ITS Directive obliges Member States to ensure that personal data processing in the ITS domain is carried out respecting fundamental rights and freedoms of individuals, with particular reference to the right to personal data protection as substantiated in the DPD (now in the GDPR) and the

---

1000 van Loenen and Grothe (n 527).
1003 ITS Directive, Art. 10(5).
ePrivacy Directive. It also encourages the use of anonymous data as a privacy protection strategy, and conditions the legitimacy of personal data processing in the ITS context to its necessity for the performance of ITS applications and services. Sectorial legislation that is incidentally related to both open data/PSI re-use and data protection maintains the same regulatory strategy used by lex generalis such as the PSI directive: ad hoc exemptions and general non-prejudice clauses.

Open data, and the infrastructure through which it is published, maintained, and re-used, are thus always meant not to be prejudicial to data protection legislation. Data protection and open data are thus explicitly, logically, and unambiguously meant to run on parallel tracks. When they happen to (possibly) intermingle, EU law provides for unequivocal exceptions and exemptions, so that their interaction can be easily subsumed under a rule. No balancing required; no proportionality involved. The legal operation to be performed is the straightforward subsumption of a case under a norm. If EU law did not provide for the exemptions on personal data protection grounds and the non-prejudice clauses with which it regulates the interaction between open data and data protection, however, a balancing exercise between countervailing rights and freedoms might be warranted. The relevant norms might be different, or new laws might be proposed. How would seeking for a proportionate balance between open data and data protection look?

5.3 Balancing as (lato sensu) Proportionality

As the aim of this thesis is to explore the prospect of balancing open data and data protection, what is left is outlining, from a normative perspective, what balancing means. The answer to the question of whether and how to balance open data and data protection depends inherently from how balancing is understood and performed. The reconciliation of competing rights and freedoms requires their comparison, and comparing completely different values has been famously described to be akin to the comparison ‘of the length of a line with the weight of a stone’:


In this respect, a first step is to clarify that – while the wording of the research questions underlying this thesis explicitly refers to the term “balancing”, it would be more precise to speak of lato sensu proportionality. As it has been put, “Early modernity […] developed two concepts in parallel: 1) Interest balancing (political and thus non-justiciable); and 2) proportionality (legal and thus justiciable)”.

1009 This research, albeit aspiring at having an interdisciplinary constituency, belongs eminently to the legal domain, and thus pursues the balance between data protection through the means of proportionality. Doing otherwise –
pursuing the balance between open data and data protection through non-justiciable interest balancing – would place the thesis in the domain of politics, rather than in the law’s.

Proportionality comes into play when a right is compressed by another right,\(^{1010}\) and entails the comparison between the interference on the right and the benefits deriving from it. “Proportionality is a fundamental concept of justice as old as organised society”.\(^{1011}\) Yet, the concept of proportionality as constructed by modern legal theory mainly arose from the European milieu – from German administrative jurisprudence, more precisely.\(^{1012}\)

From the second half of the 18th century, Prussia began to evolve from an authoritarian state, where the king’s authority and the State’s authority coincided, into a state that was governed by what English speaking countries define as the Rule of Law – the Rechtsstaat for the Germans, and the Stato di Diritto below the Alps. Proportionality emerged as complement to, and a part of, the idea of Rechtsstaat.\(^{1013}\) The Rechtsstaat principle reversed the praxis by which state action was to be regarded as legitimate even when not openly mandated by law, thus requiring an explicit textual authorization for the exercise of the State’s power.\(^{1014}\)

Proportionality, as expressed by the requirement that police powers be commensurate to the aims defined by law, complemented the Rechtsstaat principle. “The concept of the Rechtsstaat permitted the government to infringe individual rights but only when such infringement was clearly authorized by law. The principle of proportionality further limited this power, permitting the government to exercise only those measures that were necessary for achieving its legitimate goals”.\(^{1015}\) The emergence of a seminal version of the concept of proportionality is thus to be traced back to Prussian Polizeirecht,\(^{1016}\) and to the agency of administrative courts. Art. 10(2) of the 1794 Allgemeines Landrecht provided a first textual basis, embryonal but effective, for the development of the principle of proportionality by German jurisprudence.\(^{1017}\)

The constitutionalisation of the principle of proportionality had however to wait until the end of the Second World War, when the German Basic Law of 1949 established the Federal Republic and created the German Federal Constitutional Court. Reportedly,\(^{1018}\) German doctrine and jurisprudence began to argue for the recognition of proportionality as a

---

1010 Or by a public interest, in some cases.
1012 Proportionality analysis in legal reasoning has been said to have appeared much earlier: the concept traces back to ancient Greece – Engle (n 1009) 3. German jurisprudence is however generally credited with the development of proportionality in ‘modern’ legal scholarship.
1013 As reported by Cohen-Eliya and Porat, Art. 10(2) of the 1794 Allgemeines Landrecht stated that “[t]he police is to take the necessary measures for the maintenance of public peace, security and order”, and thus introduced proportionality in the Prussian legal order: Moshe Cohen-Eliya and Iddo Porat, ‘American Balancing and German Proportionality: The Historical Origins’ (2010) 8 International Journal of Constitutional Law 263, 271.
1014 Cohen-Eliya and Porat (n 1013) 271.
1015 Cohen-Eliya and Porat (n 1013) 272.
1016 Which, back then, did not identify only police powers, but “subsumed measures designed to promote the public welfare, morality, and public safety, encompassing nearly the whole of the state’s (then fairly primitive) interventions in society” - Alec Stone Sweet and Jud Mathews, ‘Proportionality Balancing and Global Constitutionalism ’ 72 98.
1017 That does not however mean that proportionality stemmed only from German polizeirecht; for instance, “[i]n international law, traces of proportionality, or necessity, date back to the very birth of international law around 1600 and has been recognised in international practice at least since the 19th century” - Christoffersen and Wahlgren (n 1011) 13.
1018 Sweet and Mathews (n 1016) 104.
constitutional principle from the very beginning. By the end of the ’50s, proportionality analysis – and the suitability/necessity/proportionality test in which it substantiates – has been solidly entrenched in German constitutionalism, and from there it spread. From German administrative and then constitutional jurisprudence, proportionality analysis has spread widely across Europe (and well beyond it), and is currently a fundamental component of several primary sources within the European multi-level fundamental rights’ framework – from individual States’ constitutional traditions to treaty-based regimes like the ECHR and the Charter.

Cohen-Eliya and Porat note that “the methodology employed by the Prussian administrative judges remained essentially formalistic and never shifted toward realism and pragmatism”. In contrast to such formal reading of proportionality, another notion was being developed in German law – the “balancing of interests”. The idea of balancing emerged as a result of radical anti-formalistic stances in German legal scholarship, the Freirechtschule. Balancing (of interests) and proportionality in German law thus derived from different legal movements with opposite perspective on legal formalism. On a similar note, the concept of ‘balancing’ – as distinct from proportionality – also characterises US doctrine and jurisprudence, where it emerged as part of a critique of legal formalism akin to the Freirechtschule’s one. In a nutshell, the idea of balancing, both in Germany and in the US, originally moved from an anti-formalist stance towards legal scholarship; proportionality, on the other hand, emerged within the frame of – and perhaps even as a consequence of – contemporary legal formalism, as a shield from the arbitrariness of power.

Indeed, another difference between proportionality and interest balancing comes to light when looking at their different finality in German constitutionalism (for the former) and in the United States’ (for the latter). The notion of proportionality emerged from German law thus derived from different legal traditions.

---

1019 Sweet and Mathews (n 1016) 104 ss.
1020 Proportionality testing “has spread across Europe, including to the post-Communist states in Central and Eastern Europe, and into Israel. It has been absorbed into Commonwealth systems – Canada, South Africa, New Zealand, and via European law, the U.K. - and it is presently making inroads into Central and South America” – Sweet and Mathews (n 1016) 74. Gunn, amongst others, reports that proportionality has been adopted as a legal instrument by extra-European common law jurisdictions like Canada or Australia as well: T Jeremy Gunn, ‘Deconstructing Proportionality in Limitations Analysis’ (2005) 19 Emory International Law Review 465, 466.
1021 For an overview on the European multi-level fundamental rights framework see Fabbrini (n 298).
1022 “(P)roportionality has also migrated to the three treaty-based regimes that have serious claims to be considered "constitutional" in some meaningful sense: the European Union (EU), the European Convention on Human Rights (ECHR), and the World Trade Organization (WTO)” – Sweet and Mathews (n 1016) 74.
1024 While the Charter dates to 2000, and it did not have binding force until the 2009 Lisbon Treaty, proportionality is a general principle of EU law, has (quasi)constitutional status by virtue of its inclusion in the Treaties, and its testing has been present in the ECJ (and then in the CJEU) jurisprudence since, to the author’s knowledge, the Judgment of the Court in Internationale Handelsgesellschaft mbH v Einfuhr- und Vorratsstelle für Getreide und Futtermittel, Case 11-70. 17 December 1970, ECLI:EU:C:1970:114.
1025 Cohen-Eliya and Porat (n 1013) 274.
1026 Cohen-Eliya and Porat (n 1013) 275.
1028 Cohen-Eliya and Porat (n 1013) 275.
administrative law in connection with the development of the Rechtsstaat, as a way to make power accountable. By contrast, in the US the notion of balancing was developed to respond to a specular issue: “the prevention of absolutism in the protection of rights”. 1029 The protection of constitutional rights in the US has a strong textual support, but, conversely, their limitation has little textual basis. Balancing, there, served as an interpretative tool that permits the compression of constitutionally protected rights by requiring their weighing against other rights and freedoms. 1030

“[I]n practice, the term "balancing" has become tantamount to the principle of proportionality”. 1031 Yet, to avoid ambiguities, it seems appropriate to specify that this thesis, being a dissertation in law and having Europe as its focus, tackles the idea of balancing data protection and open data through the (normative) lenses of proportionality, and not through the (political) ones of interest balancing. Proportionality is a legal tool, judiciable – it is not a political tool. Proportionality does not derive from anti-formalistic stances towards the law, but rather the other way around: proportionality gives formal expression to the rules to be followed when balancing countervailing rights and freedoms. Notably, while proportionality arose as a means to ensure the justiciability of conflicting rights claims, it also developed into a law-making tool, to be used by the legislature when evaluating the suitability, necessity, and proportionality of a legislative proposal, and is hence a perfect doctrinal tool to guide the evaluation of ideas about regulation, and not only of regulation per se. In other words, proportionality is not only a tool of adjudication, with which to solve legal controversies, but also a tool of creation and interpretation, a set of lenses through which to read, assess, and engineer rights and laws. 1032

5.4 LATO SENSU PROPORTIONALITY TESTING

In a generic sense, proportionality testing assesses the significance of the State’s action, 1033 e.g. the enactment of new legislation, vis-à-vis the weight of the limitations of the rights that may be compressed by it. More specifically, it is the process of either “weighing” the various interests against each other, or to find a “fair balance of interests”. Essentially, proportionality considers whether the State action is “sufficiently important” to justify the compression of a (set of) right(s). 1034 In the EU, proportionality can be found in several instruments within the European primary framework 1035 and is ubiquitous in secondary legislation both in Member

---

1029 Cohen-Eliya and Porat (n 1013) 276.
1030 Cohen-Eliya and Porat (n 1013) 276.
1032 “Proportionality is often considered in the context of limitations of rights, but the principle of proportionality is applied far beyond the confines of limitation clauses [...] In reality, the principle of proportionality is an independent means of interpretation developed alongside other canons of interpretation” - Christoffersen and Wahlgren (n 1011) 17.
1033 In a broad sense: proportionality is widely employed to solve conflicts between different rights, between a right and a public interest (constitutional law), or between a private interest and a public interest (administrative law) - Tor-Inge Harbo, ‘The Function of the Proportionality Principle in EU Law’ (2010) 16 European Law Journal 158, 158.
1034 Gunn (n 1020) 470.
1035 Most notably in Art. 5(4) TEU.
States’ law and at the EU level. The principle of proportionality is omnipresent in data protection law, particularly if read under a holistic perspective.

Proportionality is a general principle of EU law, generally meant to define the boundaries of EU powers vis-à-vis member states; it is also a core component of the European fundamental rights framework, regulating the conditions legitimising an interference with a right – data protection amongst others. In secondary data protection legislation, the principle of proportionality (as embedded, for instance, in the purpose limitation principle) has the function of a general, overarching value, which trickles down into several other provisions that require specific aspects of the processing to be suitable, necessary, and proportionate in light of the purposes set. Proportionality, both lato and stricto sensu, is data protection’s leitmotiv.

The specific wording of the provision that is used as a textual hook to justify the proportionality test (or that explicitly mandates its performance) may change, to a certain extent. For instance, Art. 8(2) ECHR mandates that “(t)here shall be no interference by a public authority with the exercise of [the right to private life] except such as is in accordance with the law”, while the Charter’s Art. 52(1) states that interferences with the rights and freedoms “must be provided for by law”. As lato sensu proportionality emerged as a judicial construct, in any case, the extent to which the specific wording impacts the execution of the test depends on the interpretation given by the relevant courts.

The provision legitimising the execution of the proportionality test may also introduce additional elements to be considered. The Charter, for instance, in its Art. 52(1), states that “(a)ny limitation on the exercise of the rights and freedoms recognised by [the] Charter must [...] respect the essence of those rights and freedoms”. The Charter thus introduces, next to the traditional lato sensu proportionality test, additional considerations about the respect of the essence of the right compressed. It is paramount to specify, however, that the place of the assessment of whether a measure is compatible with the essence of a right is next to proportionality testing, not in it. While proportionality testing comes into play to evaluate the legitimacy of interferences upon rights and freedoms, the concept of essence functions as a boundary past which an interference cannot be legitimised through its proportionality.

Proportionality testing “on the ground” has different facets depending on the subject matter at stake – e.g. data protection, freedom of expression, immigration matters, or the division

---

1036 See generally Charlotte Bagger Tranberg, ‘Proportionality and Data Protection in the Case Law of the European Court of Justice’ (2011) 1 International Data Privacy Law 239.
1037 I.e. considering also how the principle of proportionality is embedded in legal frameworks other than data protection, such as in the Treaties or in the Charter.
1038 GDPR, Art. 5(1)(b). See also Recital 4.
1039 In this case, it is safe to assume the two provisions to have identical meaning. Art. 9 and 11 of the ECHR, as opposed to Art. 8, do indeed use the wording “prescribed by law”, as the Charter does. Furthermore, the French text of the ECHR reads “prévue(s) par la loi” in Art. 9 and 11 as well as in Art. 8.
1040 See generally Lenaerts (n 924); Dawson, Lynskey and Muir (n 934).
1042 With regard to the CJEU’s jurisprudence, for instance, “the court is interpreting the proportionality principle in a variety of different ways [...] The diverging interpretations of the proportionality principle are determined
of competences between the EU and its Member States. Yet, regardless of the instrument in which it is implemented, or the specific subject matter object of the assessment, the proportionality test always has the same fundamental two-step structure. A first step is to establish the compression (or potential compression) of a right by another right. The second step, which is conditional to the infringement of the right, requires the demonstration that the encroachment pursued a legitimate aim, and that it was proportionate to such aim. This latter condition can be dubbed *lato sensu* proportionality, and is composed by three distinct and sequential sub-tests: suitability, necessity, and proportionality *stricto sensu*.

5.4.1 Suitability

The first test is suitability: the means adopted must be appropriate for furthering the goal(s) pursued by the public action considered. “There must, as it were, be a causal relationship between the measure and its object”. This sub-test is sometimes also referred to as the ‘appropriateness’ or the ‘adequacy’ test, as its aim is to assess whether the means considered are appropriate for furthering the purposes pursued.

The basic idea underlying the suitability test rests on the assumption that legislative powers must act rationally: any measure adopted must be, at least theoretically, fit to solve the issue it is meant to tackle. The measure considered must have a certain degree of efficacy and effectiveness vis-à-vis the aims it pursues, so to avoid compressing a right by virtue of a right or freedom that remains unaffected by the measure enacted.

The test considers also the legitimacy of the goals set, and some authors indeed split the consideration of the lawfulness of the compression of the right from the assessment of its suitability to achieve the goals pursued. The objectives that may justify the adoption of a measure interfering with a countervailing right are oftentimes set explicitly by the norm legitimising the proportionality balancing. For instance, the ECHR justifies a compression of the right to private life only if “in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others”. The Charter, on the other hand, allows balancing for “genuinely meet objectives of general interest recognised by

---

by the different areas of law in which it is applied, and the substance of the conflicting interests at stake” - Harbo (n 1033) 180.

1043 Cohen-Eliya and Porat (n 1013) 267.


1045 Christoffersen and Wahlgren (n 1011) 31.


1047 ECHR, Art. 8(2). Similarly, Art. 10(2) of the ECHR states that the exercise of freedom of information and expression, “since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties [...] in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary”.

172
the Union or the need to protect the rights and freedoms of others”. For the purposes of this thesis, in any case, the legality of the means is considered under the suitability sub-test.

5.4.2 NECESSITY

The second sub-test is necessity: the means must be necessary, i.e. the ones that, amongst all the suitable options available, lead to the ‘lightest’ compression of the rights and freedoms concerned. Ideally, “(t)his implies, among other things, that there must be no measure less restrictive, but adequate, available to attain the objective pursued”. Once a measure has been evaluated as being abstractly suitable for the pursuits of the objectives it sets, the following step in lato sensu proportionality testing is thus assessing whether there are measures, other than the one considered, that would entail a lesser interference with competing rights and freedoms.

The main criterion developed by doctrine and jurisprudence has traditionally been dubbed the “least restrictive means” (LRM) test: is there a suitable measure that is less intrusive upon the rights it compresses? If so, the necessity test fails; if there is no less restrictive option than the one considered, the necessity test succeeds. While the (preceding) suitability test establishes a measure’s abstract appropriateness for the achievement of the goals it pursues, the necessity test assesses the concrete extent of the compression to the rights and freedoms it leads to. The necessity test does not consider the trade-off between the benefits of a measure and the extent of its interference upon competing rights and freedoms, which is left to the subsequent stricto sensu proportionality sub-test.

However, it must be kept in mind that lato sensu proportionality testing depends on the legal context of reference, and the parameters on which the necessity sub-test is informed do too: the strict LRM test is not always the default. For instance, when it comes to the case law of the CJEU, it has been noted how the Court applies the LRM test when reviewing national legislation, and yet uses a much less strict test, the “manifest inappropriateness” of the measure considered, when reviewing EU acts. The reason can be sought in the multi-level structure of the EU, in the allocation of powers and competences between the EU and its Member states, and in the Union’s harmonization.

5.4.3 PROPORTIONALITY STRICTO SENSU

The third test is proportionality stricto sensu: the weight of the infringement of the right must be proportional to the benefits deriving from furthering the public goal. “In this sense a

---

1048 Charter, Art. 52(1).
1049 Jans (n 1044) 240.
1050 An example is given by the case CJEU, Rundfunk and Others (n 853). The CJEU, in assessing the necessity and proportionality of a measure mandating the disclosure of the amounts received as salaries or pensions of people employed by (some) public institutions for transparency purposes, while considering that the issue is for national courts to decide, also prompted the national judge to consider “whether such an objective could not have been attained equally effectively by transmitting the information as to names to the monitoring bodies alone”, and “whether it would not have been sufficient to inform the general public only of the remuneration and other financial benefits to which persons employed by the public bodies concerned have a contractual or statutory right, but not of the sums which each of them actually received during the year in question, which may depend to a varying extent on their personal and family situation”. The CJEU asked to the national court, in a nutshell, to perform a LRM test.
1051 Sauter (n 1046).
measure will be disproportionate when the restriction it causes […] is out of proportion to the intended objective or the result achieved. It could also be said that this is the proportionality principle in its true sense”.

The assessment of the stricto sensu proportionality of a measure comes after the assessment of its necessity: once it has been determined that the measure considered, amongst all the options suitable for the achievement of its aims, provides for the smallest compression of the right it interferes with, the matter becomes assessing whether the benefits to the rights and freedoms it advances surpasses the detriment to the ones it limits.

At its core, the assessment of a measure’s stricto sensu proportionality can be broken down into three subsequent steps. The first one is establishing the degree of compression of a right or freedom by the opposing right or public interest. The second step is evaluating the ‘weight’ given to the satisfaction of the conflicting right or public interest. The third step consists of establishing whether the importance of satisfying the latter justifies the infringement of the former.

There is bound to be a degree of arbitrariness inherent to this sort of assessment, and thus a number of scholars have tried to formalize the stricto sensu proportionality test by transposing its content into mathematical notation. A well-known formula used to explain stricto sensu balancing is the following: \( W_{i,j} = \frac{W_i I_i R_i^n}{W_j I_j R_j^n} \). When finding the balance \( (W_{i,j}) \) of two competing principles \( (P_i \) and \( P_j \)), one should compare their abstract weight \( (W_i \) and \( W_j \)), the concrete intensity of their interference \( (I_i \) and \( I_j \)), and the degree of empirical and normative reliability given to the consequences of the compression of one principle by another \( (R_i^n \) and \( R_i^n \), and \( R_j^n \) and \( R_j^n \) respectively).

An example might help understand how the formula represents the sort of reasoning on which a stricto sensu proportionality sub-test should be based. Let us consider the balance \( (W_{i,j}) \) between freedom of expression and information \( (P_i) \) and privacy \( (P_j) \) that must be pursued by the law and jurisprudence that ensures the possibility to report private facts about individuals that have a public role. The abstract weight of freedom of information and of privacy \( (W_i \) and \( W_j \)) is ideally the same, as both are fundamental rights in the EU. That would perhaps be different if we were to balance a fundamental right and a public interest, like national security or the prevention of crimes, or if we were to consider other legal systems (e.g., the US) where the normative underpinnings of the rights (or civil liberties) considered is starkly different.

---

1052 Jans (n 1044) 241.
1053 Klatt and Meister (n 1046) 10.
1054 See Petersen (n 1008).
1055 Most notably Alexy (n 931). Alexy’s formula is further refined in Klatt and Meister (n 1046). Comments on the formula developed by Alexy can be found e.g. in Carlos Bernal Pulido, ‘On Alexy’s Weight Formula’ in Agustín J Menéndez and Erik O Eriksen (eds), Arguing fundamental rights (Springer 2006); Lars Lindahl, ‘On Robert Alexy’s Weight Formula for Weighing and Balancing’ in Augusto Silva Dias (ed), Liber Amicorum de José de Sousa e Brito (Ediciones Almedina 2009).
1056 \( R^i \) and \( R^n \), in other words, correspond to what Alexy identifies as the “factual and legal possibility” to apply the competing principles \( P_i \) and \( P_j \) – see Alexy (n 931) 44 ss. A simplified version of the formula does not differentiate between the empirical and the normative reliability of the proportionality assessment’s premises, and reads as follows: \( W_{i,j} = \frac{W_i I_i R_i}{W_j I_j R_j} \).
1057 Again, proportionality testing is a tool and a process that can be used both by courts adjudicating controversies and by legislatures and other regulatory bodies drafting or amending laws and regulations.
The concrete intensity of the interference of the public’s freedom of information upon the privacy of individuals that have a public role (I_i), and conversely the intensity of the breach of individuals’ privacy vis-à-vis freedom of expression and information (I_j), depend on what the measure considered entails in practice. A hypothetical set of norms might shield from liability for privacy breaches any individual reporting on whatever can be tied to public persons, and the relevant jurisprudence might permit e.g. the publication of intimate pictures based on a broad conception of public interest. A different set of norms might be more conservative, preventing liability only in case the expression constituting a privacy breach is newsworthy, truthful, and exposed in a civil manner. The degree of the interference of a principle upon the other is determined, so to speak, by the nuts and bolts of the measure considered, and by its effect on the context of reference.

Yet, laws regulate the future, and the consequences of a judgement cannot be entirely foreseen when the judgement is made. Hence, the weighting of the rights and freedoms in play and the degree of interference a measure entails must also be conditioned by an evaluation of the reliability of the empirical and normative premises (R^p, and R^n). In the example above, some relevant considerations could be, for instance, about the societal value of information about public figures, or about the degree of privacy they can reasonably expect. It is not possible to predict the entire set of consequences that regulation might have, so courts and legislators (should) nuance their proportionality assessment by acknowledging the presence of ‘known unknowns’ and the possibility that ‘unknown unknowns’ exist as well.

This sort of formalisation does not, of course, reduce stricto sensu proportionality testing to a mere numerical operation. Yet, it is a useful device to describe the steps to be taken in assessing whether the detriment resulting from the compression of a right in favour of another right or public interest, by virtue of a measure that is both suitable and necessary, is also proportionate to its benefit. Again, proportionality reasoning is justiciable, and moves from theoretical stances that do not reject legal formalism, but rather embrace it. Formal logic may help us mitigate the level of subjectivity inherent to proportionality assessment, so this thesis uses Alexy’s formula for such purpose. Caveat emptor, however: legislatures and courts do not always perform the proportionality assessment as formalised above – despite proportionality’s attempt at reducing the degree of arbitrariness inherent to the exercise of power, the (qualified) interpreter of the law still enjoys a high degree of discretion.

5.5 On the Idea of Balancing Open Data and Data Protection

Let us reason, rather than de iure condito, i.e. through the lenses of positive law, de iure condendo, i.e. about how the law could be. Let us hypothesise that EU law did not provide for a set of rules exempting personal data from (open) data publishing obligations, such as Art. 1(2)(h), nor a non-prejudice clause ex Art. 1(4) of the PSI Directive, or that someone would think of proposing new legislation to ‘balance’ open data and data protection. What would it take? How would such balancing look like?


1059 See e.g. CJEU, Volker und Markus Schecke and Eifert (n 921) §71 ss, where the Court, after the suitability assessment, seems to merge necessity and stricto sensu proportionality testing.
The idea is not as far-fetched as it could seem at first glance. In the midst of the drafting of the GDPR, during the trilogue between the European Parliament, the Commission, and the Council, the latter’s general approach proposed a fairly controversial norm to balance PSI re-use and data protection. Art. 80aa of the Council’s general approach, titled “Processing of personal data and reuse of public sector information”, stated that “Personal data in public sector information held by a public authority or a public body or a private body for the performance of a task carried out in the public interest may be disclosed by the authority or body in accordance with Union law or Member State law [...] to reconcile the reuse of such official documents and public sector information with the right to the protection of personal data [...]”.

To the relief of many, the proposed Art. 80aa did not end up being adopted: as noted by the Commission during the trilogue orientation debates, it would have been in outright conflict with Art. 1(2)(h) and 1(4) of the PSI Directive. But what if Art. 80aa had been implemented into positive law? Quid iuris if we had to resolve the interaction between open data/PSI through proportionality balancing, rather than by subsuming it under a rule such as Art. 1(2)(h) and 1(4) of the PSI Directive?

This section starts to investigate the issue by considering which elements of data protection law would need to be compressed or eliminated to permit personal data to be released as open data. It then considers whether those limitations would be suitable, necessary, and stricto sensu proportionate to the aims to be achieved. The section concludes with a consideration about whether the idea of balancing open data and data protection would respect the essence of the latter.

5.5.1 WHAT WOULD IT TAKE?

Breaking down systematically, norm by norm, rule by rule, and principle by principle, all the ways open data requirements clash with data protection law, would result in an endless list of provisions. Basically, the entirety of EU data protection secondary legislation is engineered in a way that conflicts, one way or the other, with the principles on which the notion of open data is based. Is there any form of use restriction? The information considered is not open data. Access controls? Not open data.

Granted, as the previous chapter unpacked, data protection is not meant to impede personal data processing, but rather to channel it through a set of norms so to make it fair and transparent. From this perspective, data protection is not (entirely) at odds with data sharing and re-use. Yet, open data differentiates itself from less permissive modalities of information sharing and re-use by being at the extreme end of the openness spectrum — by being fully accessible, usable, and reusable, by anyone and for any purpose, in an entirely

---

1060 Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) - Preparation of a general approach. 9565/15, 11 June 2015.
1062 Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) – Partial general approach – Orientation debate. No. 16140/14, 01.12.2014, footnote 35.
1063 On the topic, see Clifford and Ausloos (n 891).
1064 De Hert and Gutwirth (n 858); De Hert and Gutwirth (n 236).
1065 See e.g. Open Data Institute (n 638).
machine-readable and open format. In this sense, the utter and unchecked openness that makes open data an autonomous concept, rather than a better sounding synonym for ‘data sharing’, is just too much to allow the processing to be channelled through the mechanics that substantiate data protection at a secondary level.

Regardless, this section aims at providing an overview of the reasons why the idea of ‘balancing’ open data and data protection would entail, provided that the term ‘open data’ is not diluted, an extreme compression of the right to personal data protection. Many of the checks and balances of which data protection law is made are in direct opposition to the requirements underlying open data. Within the boundaries of the current legal framework, ultimately, information can either be personal data or open data.

5.5.1.1 Data Protection’s Far-Reaching Scope

A preliminary consideration concerns the width of the scope of EU data protection law. The material, territorial, and personal scope of the GDPR (and of the DPD) have often been regarded as overly extensive; while that may be a matter of opinion, there is no denying that EU data protection is quite far-reaching. As a consequence, any shift in the balance between open data and data protection would impact a very wide range of processing operations, run by many different data controllers, inside and outside of the EU.

As for the personal scope, the GDPR defines as data controller “the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data”. Any natural or legal person can qualify as controller. Controllership is factual, in that it is determined from the concrete behaviour of the parties involved in the processing, and substantial, since it does not require nor is influenced by any kind of formal stipulation or agreement, but only by the possibility to determine the purpose and means of processing. The determination of the purpose of processing is a functional test, aiming at the clear allocation of the processing’s responsibility: “determining” something is having factual control over it, stemming from an explicit or implicit legal competence, or by mere factual influence. The width of the notion of (joint) data

---

1066 See e.g. Purtova, ‘The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law’ (n 903); Brkan, ‘The Unstoppable Expansion of the EU Fundamental Right to Data Protection: Little Shop of Horrors?’ (n 903); van Loenen, Kulk and Ploeger (n 977).

1067 GDPR, Art. 4(7). If the purposes and means of processing are determined by EU or MS law, the controller or the specific criteria for determining controllership can be set by law too. Controllership can also be joint, where two or more controllers jointly determine the purposes and means of processing (see GDPR, Art. 26). Processors process personal data on behalf of the controller (GDPR, Art. 4(8)). They are separate legal entities from the controller, tasked with processing data on its behalf. The qualification of processor, like the one of controller, is factual: it derives from the activities concretely performed. Processors derive their legitimacy from the tasks assigned to them by the controller, and cannot process data at their discretion, at least not without becoming controllers themselves. Every other “natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorised to process personal data” is a third party (GDPR, Art. 4(10)).

controller has been confirmed – perhaps outright extended – by the jurisprudence of the CJEU.\footnote{1069}

As for its territorial scope,\footnote{1070} the GDPR applies to processing performed in the context of the activities of an establishment of a controller or a processor in the EU, regardless of whether the processing takes place in the Union or not. When the processing involves data subjects that are in the EU, the GDPR also applies to the activities of controllers and processors that are not established in the EU, if they are offering goods or services to EU data subjects, or monitoring their behaviour.\footnote{1071} The GDPR can also apply if the controller, despite not being established in the EU, is located in a place where MS law applies by public international law.\footnote{1072} Aside from the extensive territorial scope of the GDPR, the long-arm reach of EU data protection law is granted by its provisions on transborder data flows –\footnote{1073} the rules legitimising transfers of personal data to third countries or to international organisations. Finally, in addition to the territorial reach of the GDPR, EU data protection law has had a strong shaping effect on the laws of countries that are not subject to the EU’s jurisdiction,\footnote{1074} and on the practices of multinational organisations.\footnote{1075}

The GDPR’s material scope is triggered by the processing of personal data, wholly or partly by automated means, and other than by automated means if the personal data are meant to form (part of) a filing system.\footnote{1076} Processing is described, extremely broadly, as “any operation or


\footnote{1070}GDPR, Art. 3. See EDPB, ‘Guidelines 3/2018 on the Territorial Scope of the GDPR (Article 3) - Version for Public Consultation’ (2018). The GDPR hybridises the territoriality principle – according to which the relevant factor is the location of the actors, and to which the DPD was informed – with the protection principle, where the connecting factor is the location of the data subjects, and which emphasizes the actions of the controller. At first glance, the territorial scope of the GDPR appears to be much broader than the previous DPD’s reach. The latter was applicable to the processing in the context of the activities of an establishment of the controller on the territory of the MS; if the controller is not established on the MS’s territory, but in a place where its national law applies by virtue of international public law; and if the controller is not established on EU territory but, for purposes of processing personal data, makes use of equipment situated in a MS, unless only for purposes of transit through EU territory. The DPD’s scope has oftentimes been interpreted extensively by the CJEU, and so the scale of the extension of the territorial scope is thus concretely smaller than what suggested by a literal comparison between the wording of Art. 3 of the GDPR and Art. 4 of the DPD. See Lokke Moerel, ‘Back to Basics: When Does EU Data Protection Law Apply?’ (2011) 1 International Data Privacy Law 92; Lokke Moerel, ‘The Long Arm of EU Data Protection Law: Does the Data Protection Directive Apply to Processing of Personal Data of EU Citizens by Websites Worldwide?’ (2011) 1 International Data Privacy Law 28.

\footnote{1071}As long as it takes place in the EU.

\footnote{1072}GDPR, Art. 3(3)

\footnote{1073}GDPR, Art. 44 ss. See generally Christopher Kuner, Transborder Data Flows and Data Privacy Law (Oxford University Press Oxford 2013).

\footnote{1074}See e.g. Birnhack (n 66); Greenleaf (n 66); Graham Greenleaf, Asian Data Privacy Laws: Trade & Human Rights Perspectives (OUP Oxford 2014).

\footnote{1075}See Bradford (n 742).

\footnote{1076}The only type of personal data processing falling out is performed entirely by non-automated means on data which is not meant to be part of a filing system, defined as “any structured set of personal data which are accessible according to specific criteria, whether centralised, decentralised or dispersed on a functional or geographical basis” (GDPR, Art. 4(6)). The GDPR does not however apply to the processing of personal data during an activity outside the scope of EU law; by Member States when carrying out activities which fall within the scope of Chapter 2 of Title V of the TEU; by a natural person during a purely personal or household activity; by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the prevention of threats to public security.
set of operations which is performed on personal data or on sets of personal data, whether or not by automated means”. Any action performed upon personal data in whatsoever fashion constitutes processing. Personal data is “any information relating to an identified or identifiable natural person […] one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person”. According to Recital 26 of the GDPR, identifiability must be constructed taking into account all the means reasonably likely to be used, by the data controller or by another person, in order to identify the data subject. The breadth of the material scope of EU data protection law has been confirmed by the CJEU in several occasions.

The GDPR’s personal, territorial, and material scopes have thus been drafted to be very broad, and the width of the textual provisions has often been either established or expanded by the interpretative activity of the CJEU. As a consequence, any attempt at balancing it with the right to re-use PSI, let alone as actual open data, would have to be mindful of the scale at which EU data protection operates, and hence of the ‘ripple effects’ entailed by such balancing attempt. In other words, the sheer range of persons and kinds information to which EU data protection law is applicable, and the gamut of territories over which it has a direct or indirect effect, are bound to have a considerable impact on the weighing exercised involved in balancing open data and data protection.

5.5.1.2 A MATTER OF PRINCIPLES

Personal data processing must be performed following a number of principles, enumerated by Art. 5 of the GDPR. The list results from the evolution of international data protection legislation, and its core can be tracked back to the Fair Information Principles (FIPs, also referred to as Fair Information Practice Principles, or FIPPs) that resulted from the first, seminal solutions proposed for the “computers and privacy” issue since the ’70s. The principles ex Art. 5 GDPR are present across privacy and data protection legislation all over the world, so

---

1077 GDPR, Art. 4 (2).
1078 Information is constructed as a very broad concept, encompassing whatever is susceptible to be represented through data: names and last names, ID numbers, license plates, IP addresses, fingerprints and other biometrics, for instance. It may be argued that the concept of information in EU law has been, so far, left as unspecified (and underdeveloped) as possible – see Bygrave, ‘Information Concepts in Law: Generic Dreams and Definitional Daylight’ (n 852).
1079 The Article 29 Working Party clarifies the “relation criterion” by stating that “to consider that the data “relate” to an individual, a "content" element OR a "purpose" element OR a "result" element should be present”. The “content” element is present when the data is about a specific person in the biographical sense of the term. The “purpose” element is present where the data is used (or is likely to be used) to achieve a particular goal in relation to an individual, i.e. where its purpose involves a specific natural person. The “result” element is present where, regardless of the information’s content and of the processing’s purpose, the concrete data processing instance is (likely) to result in an impact on a specific individual’s rights and interests. Those conditions are to be seen as alternative, rather than cumulative – one suffices. See Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40).
1080 GDPR, Art. 4 (1).
1081 See e.g. CJEU, Patrick Breyer V Bundesrepublik Deutschland (n 845). See also CJEU, Peter Nowak v Data Protection Commissioner (n 845); CJEU, YS and others (n 845).
1082 See the previous chapter. See generally Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19); Bygrave, Data Privacy Law: An International Perspective (n 269); Gellman (n 295).
much that they can be described as the *lingua franca* of global information privacy and data protection. Even a cursory comparison between the principles ex Art. 5 GDPR and the requirements defining the concept of open data reveals the basic, fundamental incompatibility between the two.

Personal data must be processed lawfully, fairly, and in a transparent manner, following the lawfulness, fairness, and transparency principles. The substantial content of the fairness principle is quite elusive, and the terms ‘fair’ and ‘fairness’ are found in connection with the concepts of lawfulness and transparency of the processing throughout the entire GDPR. The transparency principle mandates that natural persons are to be made aware that personal data concerning them are collected and used, and of the extent to which their processing takes place; it requires that any information relating to personal data processing must be easily accessible and understandable, made available in clear and plain language. Last, the lawfulness principle states the conditions under which the personal data processing can happen legally. Those conditions – the grounds for lawful processing – are listed, exhaustively, in Art. 6 of the GDPR.

For the processing to be legal, at least one of the following conditions must be satisfied: (1) the data subject has given consent to the processing; (2) processing is necessary for the performance or stipulation of a contract to which the data subject is party, or (3) for compliance with a legal obligation to which the controller is subject, (4) to protect the vital interests of the data subject or of another natural person, (5) for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller; or (6) for the purposes of the legitimate interests pursued by the controller or by a third party, except where overridden by the interests, rights, or freedoms of the data subject.

The fact that personal data processing, to be lawful, must be explicitly based on one of the grounds exhaustively listed in Art. 6 is inherently at odds with the aim underlying open data – making information usable and reusable by anyone and for any purpose. The grounds amount to the need for a legitimate overarching reason justifying the processing of personal data, while the rationale underlying the requirements in which the concept of open data substantiates

---

1083 Hartzog (n 231) 59.
1084 GDPR, Art. 5(1)(a).
1085 See Clifford and Ausloos (n 891).
1086 See GDPR, Recital 39.
1087 Consent is defined by Art. 4(11) of the GDPR as “any freely given, specific, informed and unambiguous indication of the data subject’s wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data”. Further conditions for consent are specified in Art. 7 and 8 of the GDPR. On consent in EU data protection law, see generally Eleni Kosta, *Consent in European Data Protection Law* (Martinus Nijhoff Publishers 2013).
1088 In this case, the legal basis for processing must be laid down by EU or MS law to which the controller is subject: GDPR, Art. 5(3). The purpose of the processing shall then be determined by that legal basis.
1089 Here, too, the basis for processing must be laid down by EU or MS law to which the controller is subject: GDPR, Art. 5(3). The purpose of the processing shall then be determined by that legal basis, and be necessary for the performance of the task carried out in the public interest, or in the exercise of the controller’s official authority.
1090 The legitimate interest ground does not however apply to processing carried out by public authorities in the performance of their tasks: GDPR, Art. 6(1). See generally Article 29 Data Protection Working Party, ‘Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller under Article 7 of Directive 95/46/EC’ (2014).
rejects gatekeeping mechanisms such as the need for a ground ex Art. 6 GDPR. Both the EDPS and national data protection authorities\textsuperscript{1091} have often stressed the need for PSI releases to be compliant with the lawfulness principle, and hence covered by at least one of the grounds ex Art. 6 of the GDPR. If one subsumes the interaction between PSI and data protection under Art. 1(2)(h) and 1(4) of the recast of the PSI Directive, as positive law would require, it results that the necessity for PSI releases to be compliant with data protection law (1) derives naturally from the hierarchical relationship between PSI and data protection, and (2) is not inherently an obstacle for personal data to be released as PSI. It is questionable, however, whether such PSI could still fall under a strict definition of open data, rather than under a less permissive option on the openness spectrum.

Personal data must also be collected for specified, explicit and legitimate purposes, without being further processed in a manner that is incompatible with those purposes\textsuperscript{1092} (purpose limitation principle).\textsuperscript{1093} As it has been oftentimes stressed, the purpose limitation principle is one of the biggest hurdles in reconciling the rationale underlying open data with the principles substantiating the right to personal data protection.\textsuperscript{1094} At the same time, if there were a single principle to be picked as a banner for the ways in which the right to data protection is given substance, purpose limitation would probably be it. The principle can be split into two main elements: purpose specification, according to which the data controller must only collect personal data for specified,\textsuperscript{1095} explicit,\textsuperscript{1096} and legitimate\textsuperscript{1097} purposes, and compatible use, according to which collected data must not be further processed in a way that is incompatible\textsuperscript{1098} with the collection’s purposes.\textsuperscript{1099} Both parts are in fundamental conflict with the ethos underlying open data. Purpose specification clashes with the idea of sharing information for the sake of sharing it, hoping that it will eventually be used to engender transparency, efficiency, and growth; compatible use is at odds with the concept of ‘serendipitous reuse’, the idea for which processing data for processing’s sake leads to spontaneous and unexpected generation of new knowledge and (hence) value.

\begin{footnotes}
\item[1091] See European Data Protection Supervisor (n 991); Article 29 Data Protection Working Party, ‘Opinion 06/2013 on Open Data and Public Sector Information (‘PSI’) Reuse WP207’ (n 521).
\item[1092] Further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes is however not considered to be incompatible with the initial purposes: GDPR, Art. 5(1)(b).
\item[1093] GDPR, Art. 5(1)(b).
\item[1094] See Tal Z Zarsky, ‘Incompatible: The GDPR in the Age of Big Data’ (2016) 47 Seton Hall L. Rev. 995. Zarsky discusses big data, rather than open data, but the considerations about purpose specification and limitation are applicable, \textit{mutatis mutandis}, to the latter as well.
\item[1095] The purposes of the processing must be specified beforehand, at the latest when the collection takes place, and must be clearly and specifically identified: overly generic purposes are unlikely to meet the specification threshold. See Article 29 Data Protection Working Party, ‘Opinion 03/2013 on Purpose Limitation’ (n 533).
\item[1096] I.e. unambiguous and precisely stated; in other words, implied or secret purposes are not allowed. See Article 29 Data Protection Working Party, ‘Opinion 03/2013 on Purpose Limitation’ (n 533).
\item[1097] Legitimacy, in the context of the purpose limitation principle, can be read as accordance with the law in its broadest sense.
\item[1098] Any processing operation following the collection must be considered ‘further processing’ and must thus meet the requirement of non-incompatibility. Incompatibility derives from a substantial case-by-case assessment of the circumstances pertaining to the processing, \textit{informed inter alia} on the relationship between the collection’s purposes and the further processing’s one, on the context of the processing and on the data subject’s reasonable expectations about it, on the nature of the processing and on its expected impact, and on the safeguards foreseen to ensure fair processing.
\item[1099] Article 29 Data Protection Working Party, ‘Opinion 03/2013 on Purpose Limitation’ (n 533) 11.
\end{footnotes}
The data minimisation principle,1100 which mandates that personal data must be adequate, relevant, and limited to what is necessary, would intuitively seem another major point of conflict between data protection and the values and principles underlying open data. Yet, minimization – the limitation of the processing by what is necessary – has to be considered in relation to the processing’s purposes, so that their specification effectively functions as a sort of benchmark to determine what ‘minimized’ means. It is not untrue that the data minimization principle is inherently at odds with the high-volume data processing on which open data (and other connected notions, like big data and machine learning) are based.1101 Yet, the possibility for data controllers to tune the extent to which the processing has to be minimized by virtue of their control over its purposes allows for a considerable degree of leeway, and at the same time links the purpose limitation principle and the data minimization one in an inextricable fashion. In other words, data minimization is in overt contrast with the idea that “the more (information), the merrier” on which the open data narrative is centred, and yet, to a large extent, such contrast derives indirectly from the purpose limitation principle, rather than from data minimization per se.

The data must also be accurate and kept up to date; personal data that are inaccurate, having regard to the purposes for which they are processed, must be erased or rectified without delay (accuracy principle),1102 if it is reasonably feasible to do so. The accuracy principle would intuitively be compatible with the principles and goals underlying open data. Data quality is indeed a prominent issue in data sharing and re-use, as much so as in data protection; the underlying idea is that good data produces good (as in ‘accurate’) outcomes, and that, on the other side of the coin, “garbage in, garbage out”. That may very well be true, in principle, but the function of the accuracy principle in data protection is sheltering data subjects from the harm that may derive from the processing of outdated and inaccurate information about themselves, not ensuring that data controllers are provided with quality data to process. The apparent convergence between the accuracy principle and the principles underlying open data is given by an identity of means – accuracy to foster data quality – but is held back by a divergence of ends: the idea that a good input leads to a better output, on one hand, and the will to prevent damage to data subjects resulting from outdated or inaccurate personal data processing, on the other.

Personal data, moreover, following the storage limitation principle,1103 must be kept in a form that allows the identification of the data subjects for no longer than necessary for the processing’s purposes. It may, nevertheless, still be stored for longer periods, if processed solely for archiving purposes in the public interest, for scientific or historical research, or for statistical purposes. Yet, again – which are the purposes underlying open data sharing? Re-use of open data has generally1104 purposes that are specific enough to be specifiable ex ante, but

---

1100 GDPR, Art. 5(1)(c). See also GDPR, Recital 39: “Personal data should be processed only if the purpose of the processing could not reasonably be fulfilled by other means”.
1101 See Zarsky (n 1094).
1102 GDPR, Art. 5(1)(d).
1103 GDPR, Art. 5(1)(e). “Personal data should be adequate, relevant and limited to what is necessary for the purposes for which they are processed. This requires, in particular, ensuring that the period for which the personal data are stored is limited to a strict minimum”: GDPR, Recital 39.
1104 Big data analytics, where information is processed to determine interesting patterns and correlations, rather than to confirm or falsify a given hypothesis, might differ from the standard – see e.g. Mayer-Schönberger and Cukier (n 13); Kitchin, The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences (n 4).
the publication of open data often happens without any apparent purpose, or with the all-purpose, generic aim of stimulating transparency, efficiency, and growth. The application of the storage limitation principle to open data sharing and re-use, currently, would entail the anonymization of the information collected and processed, preliminary to its release as a different anonymised and derivative data set. How would the storage limitation principle fare if it had to be ‘balanced’ with open data principles in another way?

Both the accuracy and the storage limitation principles highlight another factor that is paramount for the relationship between open data initiatives and personal data protection: the nature of information as a good. Information is a peculiar kind of good, in that it is non-rivalrous (i.e. its enjoyment by a person does not preclude its consumption by another person) and non-excludable (i.e. it is not normally possible to exclude others from its use); digital information, in particular, is also reproducible at a marginal cost that is often close to zero. The economic characteristics of digital information are at the basis of the open data creed and of European PSI legislation, which rest on the idea that the reuse of old information for new purposes, unforeseen at the time of the data collection, would lead to newfound knowledge and growth. Those characteristics also make it so that data protection principles such as accuracy and storage limitation, which do not necessarily conflict with the principles underlying open data, end up being devoid of their meaning by virtue of the sheer reproducibility of digital information. How can accuracy and storage limitation be effectively pursued when the information they mean to channel is specifically engineered to be duplicated and disseminated as easily as possible?

The integrity and confidentiality principles mandate that, when processing personal data, an appropriate level of security (e.g. protection against unauthorised or unlawful processing, and against accidental loss destruction or damage) must be ensured through the implementation of appropriate technical and organisational measures. The publication of information as open data is fundamentally at odds with the confidentiality aspect of the security principle. Granted, security is not tantamount to mere confidentiality – it also requires ensuring the data’s integrity and guaranteeing its availability; from this perspective, there are aspects of security that are not in overt conflict with the requirements on which open data is based. Yet, information cannot be called open data if the possibility to access or use it is restricted, which rules out the implementation of any sort of (confidentiality-related) security measure successive to the publication. Arguably, the only ways to ‘reconcile’ confidentiality and open data is to anonymise the information before its publication, as it is current standard practice, and which would take the resulting data out of the material scope of EU data protection law altogether, or to pseudonymise it. Any other technical and

---

1105 See e.g. Zuiderveen Borgesius, Van Eechoud and Gray (n 41); Scassa (n 614).
1106 Although technical protection measures (e.g. a Digital Rights Management system, or DRM) can sometimes be implemented, information remains a non-excludable good by default.
1108 GDPR, Art. 5(1)(f).
1110 I.e. processing of personal so that it “can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical
organisational measure amounts to the kind of access control or use restriction that is, by definition, incompatible with the concept of open data.

Controllers, finally, by virtue of the accountability principle, are responsible for, and must be able to demonstrate compliance with, the principles enumerated in Art. 5(1). The accountability principle would not intuitively appear to be in fundamental conflict with open data. Yet, practically, many of the mechanisms through which accountability can be implemented in data protection law – access controls, logging, and record-keeping – are hard to reconcile with the extreme accessibility and usability underlying open data. Even the practice of requiring users’ registration before downloading an open data set from an open data portal with the purpose of studying who re-uses such information (aiming at anticipating which future data release to prioritise on account of the re-users’ demography, for instance) can be seen as a barrier to accessibility, and thus deprive the data of its ‘open’ connotation.

The core principles in which European data protection law is substantiated are thus incompatible with the requirements underlying open data. Granted, EU data protection is not a prohibitive regime – the processing of personal data is generally allowed, if and to the extent that it complies with the applicable norms – and thus, from this perspective, it is not necessarily incompatible with (personal) data sharing and re-use. Yet, the all-purpose (or purposeless) publication of information and the utterly unconstrained possibility to access and re-use it, which are foundational to the concept of open data, outstretch the permissive nature of personal data protection far beyond its breaking point. From this perspective, open data tenets, if fulfilled in their entirety, reject the very core of the notions of balancing and proportionality that are bound to come into play when two countervailing rights clash.

5.5.1.3 Diluting rights

The norms contained in Art. 5 are principles, rather than rules: optimization requirements mandating that the processing must be as transparent and constrained as possible within a given case, as opposed to rules, which are either complied with or not. Yet, those principles permeate the whole European data protection framework, detailing themselves across a multiplicity of more specific provisions and often substantiating themselves in other norms that do function as rules, rather than principles. The integrity and confidentiality principles, for instance, instantiate in the provisions about data security; the purpose limitation principle and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person: GDPR, Art. 4(5). Pseudonymous information is still personal data under the GDPR: see Recital 26.


1112 The mere number of ‘hits’ (e.g. the number of GET requests for a particular web page) does not say much about which users download what (and from where), and does not provide for a particularly useful metric to determine users’ preferences and behaviour.

1113 See Fuster and Gutwirth (n 882). See also De Hert and Gutwirth (n 858); De Hert and Gutwirth (n 236).

1114 See Lynskey, The Foundations of EU Data Protection Law (n 19) 30; Fuster and Gutwirth (n 882).

1115 See Alexy (n 931).

1116 GDPR, Art. 32 ss.
is embodied by the plethora of provisions that tie the rules applicable to the processing activities to their purpose. 1117 Data subjects have several rights, and controllers a corresponding set of obligations, whose fulfilment would either be hard to guarantee or rendered practically meaningless if and when personal information is shared as open data. The non-rivalrous and non-excludable nature of information as a good, coupled with the unfettered availability and reusability necessary to qualify data as ‘open’, makes it so that – even when it would be theoretically possible to nominally guarantee data subjects’ rights and to comply with the controllers’ obligations in a given processing instance – it would not really lead to the high level of safeguard that EU data protection law aims at guaranteeing.1118

Data subjects have the right to be informed about the processing of their personal data,1119 whether provided directly by them or not, in a concise, transparent, intelligible, and easily accessible form, in a clear and plain language. Data subjects have also the right to access the data processed by the controller, after obtaining confirmation, if necessary, as to whether or not personal data concerning him or her are being processed.1120 The information to be provided to data subjects concerns, broadly speaking, details about the data controller, about the processing operations performed, and about the rights available to the data subject. Its purpose is to curb the information asymmetry between controllers and data subjects that is inherent to personal data processing, and to permit and facilitate the exercise of the rights available to the latter. Releasing personal data as open data would require a partial relaxation of the detail of the information to be provided by the entity releasing the data. Much of it pertains to the purposes of the processing performed, and the idea of ‘serendipitous reuse’ that underlies the open data ethos is at odds with the specification of (and of any limitation to) the processing’s purposes. The issue does not seem to exist for the entity re-using such data, as that would be a different processing instance with its own purposes, which can (and should) be specified ex ante and communicated timely to data subjects.

The fact that open data re-use does not involve direct collection from data subjects, in conjunction with the fact that open data sets are likely to contain personal data about a multiplicity of different persons, might render the provision of information practically unfeasible, or demand a disproportionate amount of time and resources. Yet, in such cases, where the data is not collected directly from data subjects, and informing them would be practically impossible or require disproportionate effort, Art. 14(5)(b) of the GDPR already allows to derogate to the obligations to inform them.1121 While the derogation ex Art. 14(5)(b) should be interpreted narrowly,1122 it may still be used to ease the burden upon controllers

1117 See e.g. Art. 11, 25, and 35, amongst many. See generally Article 29 Data Protection Working Party, ‘Opinion 03/2013 on Purpose Limitation’ (n 533).
1118 Cfr. CJEU, Heinz Huber v Bundesrepublik Deutschland, Case C-524/06, 16 December 2008, ECLI:EU:C:2008:724, §50; Maximilian Schrems v Data Protection Commissioner (n 915) §39; Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, C-131/12, 13 May 2014, ECLI:EU:C:2014:317, §38 and the case law cited therein.
1119 GDPR, Art. 12 ss.
1120 GDPR, Art. 15.
1121 In particular for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, sectors which are fairly close to the open data milieu. Art. 14(5)(b) specifies that such derogation is still subject to the conditions and safeguards ex Article 89(1) of the GDPR. In any case, when making use of the exception ex Art. 14(5)(b), controllers must take appropriate measures, e.g. making the information in question publicly available.
1122 As all derogations and limitations – cfr. e.g. CJEU, Tietosuojavaltuutettu v Satakunnan Markkinapörssii Oy, Satamedia Oy, Case C 73/07, 16 December 2008, ECLI:EU:C:2008:727, §56; Volker und Markus Schecke and Efert
reusing datasets that contain the personal data of a multiplicity of data subjects which have not been obtained directly from them. In a way, Art. 14(5)(b) of the GDPR already takes partial care of the issues concerning the information to be provided to data subjects that may arise when thinking about releasing personal data as open data. However, it does so by shifting the burden on data subjects, and it would need to be interpreted and used more extensively than currently allowed if personal data were to be released as open data.

Data subjects also have the right to rectify inaccurate personal data about them,\textsuperscript{1123} and to erase them\textsuperscript{1124} if certain conditions are met. Data subjects, furthermore, can restrict the processing of their personal data,\textsuperscript{1125} or outright object to it.\textsuperscript{1126} While in principle open data sharing (and its subsequent re-use) does not impede the exercise of those rights, it would still practically dilute their effectiveness and capacity for protection. A data subject may very well request the rectification, restriction, or erasure of his or her personal data to the entity publishing it, or object to its processing. Yet, all ‘downstream’ re-users who downloaded the (static) dataset before the controller fulfils the data subject’s request would still be processing his or her personal information, and the data subject would thus need to reiterate his or her request to each of them.

The data protection issues that may derive from the duplication of data are of course not specific to open data re-use, but rather derive from the non-rivalrous nature of information as a good. The GDPR, in its Art. 19, already tackles the issue by obliging the controller to communicate, to each recipient to whom the personal data have been disclosed, any rectification or erasure of personal data, or restriction of its processing.\textsuperscript{1127} Art. 19 of the GDPR also requires the controller to inform the data subject, upon his or her request, about the identity of those recipients. Akin to the obligation to inform data subjects in case the data is not collected directly from them ex Art. 14 of the GDPR, controllers can be exempted from the obligation to notify the recipients of the personal data whose processing has been restricted, or that have been rectified or erased, if that would prove impossible or involve a disproportionate amount of effort,\textsuperscript{1128} for example when the notification obligation concerns a dataset containing pseudonymised individual records relating to a very large number of natural persons that are indirectly identifiable by an entity other than the data controller.

Open data policies and practices are purposely structured to enhance data sharing and re-use as much as possible, and the mechanisms with which data controllers can ascertain and record who uses the information they publish – logging, registration, and the likes – are tendentially incompatible with the permissionless, unmediated nature of open data. As a consequence of the fact that, simply put, open data publishers do not really know who re-users are, it is likely that the exemption contained in Art. 19 of the GDPR would have to be shifted from being an exception to being the rule.

\textsuperscript{1123} GDPR, Art. 16.
\textsuperscript{1124} GDPR, Art. 17.
\textsuperscript{1125} GDPR, Art. 18.
\textsuperscript{1126} GDPR, Art. 21.
\textsuperscript{1127} Notably, the notification obligation ex Art. 19 of the GDPR does not apply to the right to object.

\textsuperscript{1128} GDPR, Art. 19.
There are a number of other rights granted by the GDPR that would not seem to be directly impacted by the pursuit of a balance between open data and data protection. On the other side of the coin, data controllers (and processors) are also subject to a set of obligations which goes beyond the respect of the rights and principles discussed above, and that can be hardly complied with if personal data were to be shared and re-used as open data.

5.5.1.4 WEAKENING OBLIGATIONS

Data controllers, in compliance with the accountability principle, are bound to respect and apply the principles ex Art. 5(1), and to enable the exercise of the data subject rights ex Art. 12-22. In that respect, the sections above highlighted (some of) the issues that would arise if personal data were to be shared as open data, and that would thus need to be ironed out if data protection law were to be balanced with the open data tenets. Data controllers also have a number of additional obligations (mainly revolving around logging, record-keeping, and information and systems security) that cannot really be complied with in an open data setting – or at least, not in their entirety.

In general, controllers must implement appropriate technical and organisational measures to ensure and demonstrate that processing is performed in compliance with the GDPR, and maintain a written record of the processing activities they perform. Such record must contain, amongst other kinds of material, information about the purposes of the processing, the categories of data subjects involved, of the personal data processed, and of their recipients. The record must also include information about eventual personal data transfers to third countries or international organisations, the envisioned time limits for erasure, and a general description of the security measures adopted by the controller. Likewise, each processor must maintain a record of the processing activities carried out on behalf of the controller; such record must contain, inter alia, the categories of processing performed, an indication of the eventual personal data transferred to third countries or international organisations, and where possible a general description of the technical and organisational security measures adopted.

Re-users do not have particular reasons not to comply with such obligation – personal data released as open data by another data controller would be a source like any other to them –

1129 The GDPR grants data subjects the right to data portability as well, allowing them to receive their personal data, in a structured, commonly used, and machine-readable format, and to transmit them to another controller without hindrance (GDPR, Art. 20) – on the subject, see De Hert and others (n 869); Ursic (n 869). Data subjects have also the right not to be subject to automated decision-making, e.g. profiling (GDPR, Art. 22), which comes, to some extent, with a ‘right to an explanation’ – see Andrew D Selbst and Julia Powles, ‘Meaningful Information and the Right to Explanation’ (2017) 7 International Data Privacy Law 233; Edwards and Veale (n 818); Sandra Wachter, Brent Mittelstadt and Luciano Floridi, ‘Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation’ (2017) 7 International Data Privacy Law 76. Data subject have, finally, the right to lodge a complaint with an administrative supervisory authority, to a judicial remedy against both a supervisory authority and the data controller or processor, and to obtain compensation for both material and non-material damages) – GDPR, Art. 77 ss; on the resulting fines, see Paul Nemitz, ‘Fines under the GDPR’ in Ronald Leenes and others (eds), Data Protection and Privacy - The Internet of Bodies (Hart Publishing 2018).

1130 GDPR, Art. 5(2).

1131 GDPR, Art. 24. Compliance with the GDPR and adherence with its principle and with the state of the art on the matter can be demonstrated by the adherence to recognised codes of conduct, standards, seals, and certifications (GDPR, Art. 24(3)) – see Art. 40 ss.

1132 GDPR, Art. 30.

1133 GDPR, Art. 30(2).
but publishers would have significant issues in recording the information required ex Art. 30 of the GDPR if the personal data in consideration were to be released as actual open data. The purposes of the processing, yet again, are undefined by default in the open data ethos; the recipients cannot be ascertained without implementing access control measures that are seen as a limit to re-use; security measures aiming at ensuring the information’s confidentiality cannot be adopted by definition. The obligation to maintain a record does not apply to organisations with fewer than 250 persons – unless the processing is likely to result in a risk to the rights and freedoms of data subjects, is not occasional, or it includes personal data that is sensitive or relates to criminal convictions and offences. Such exemption might apply to certain categories of re-users – ones that employ fewer than 250 individuals and that process occasionally low-risk personal data – but not to all of them, and it seems unlikely that it would affect the kind of entity that publishes open data.

The controller must also integrate the general data-protection principles and the necessary safeguards into the processing itself (Data Protection by Design), and implement appropriate technical and organisational measures for ensuring that – by default – only personal data which are necessary for each specific purpose are processed (Data Protection by Default). As for the latter, the Art. 25(2) of the GDPR makes specific mention of the fact that those measures shall ensure that, by default, personal data are not made accessible, let alone re-usable, without the individual’s intervention, to an indefinite number of persons – which, in a way, is open data’s whole point. From this perspective, open data, and its “open by default” maxim, are the polar opposite of Data Protection by Default. Likewise, open data sharing and re-use, to a large extent, are conflicting with the confidentiality-related aspects of data security, which is closely connected with the Data Protection by Design requirements.

Data controllers have, indeed, the responsibility to ensure the security of the processing and of the personal data involved – rectius, a level of security appropriate to the risk foreseen, considering the state of the art, the costs of implementation and the nature, scope, context and purposes of processing. Data security, i.e. “the availability, authenticity, integrity and confidentiality of stored or transmitted personal data”, as the term should suggest, is an

1134 GDPR, Art. 30(5).
1135 GDPR, Art. 25(1). See generally Rubinstein (n 275); Cavoukian (n 274); Ann Cavoukian, ‘Privacy by Design: Leadership, Methods, and Results’ in Serge Gutwirth and others (eds), European Data Protection: Coming of Age (Springer Netherlands 2013); Hartzog (n 231). On Privacy by Design and by Default in the GDPR, specifically, see Lee A Bygrave, ‘Data Protection by Design and by Default: Deciphering the EU’s Legislative Requirements’ (2017) 4 Oslo Law Review 105. See also, more practically, Danezis and others (n 306).
1136 GDPR, Art. 25(2). See Bygrave, ‘Data Protection by Design and by Default: Deciphering the EU’s Legislative Requirements’ (n 1135).
1137 GDPR, Art. 32. Notably, the influence of the risk-based approach permeates the dispositions about the technical and organisational norms to be observed by controllers, sanctioning that technical and organisational security measures shall be implemented (and, conversely, evaluated) considering the state of the art, the cost of implementation, and the nature, scope, context, and purposes of processing, as well as the risks (of varying likelihood and impact) for the rights and freedoms of natural persons involved - see e.g. GDPR, Art. 24, 32, 35. Data controllers must also notify, without undue delay, eventual personal data breaches to the competent supervisory authorities, and to data subjects too, where the breach is likely to present high risks to their rights and freedoms (GDPR, Art. 33, 34). The concept of personal data breach in EU data protection law is fairly extensive, as it means “a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed”: GDPR, Art. 4(12). See Maria Karyda and Lilian Mitrou, ‘Data Breach Notification: Issues and Challenges for Security Management’, MCIS (2016).
1138 GDPR, Recital 49.
integral part of data protection. Open data sharing and re-use, however, are quite hard to reconcile with the implementation of many of the security measures relevant for data protection. Measures aiming at ensuring the availability of personal data go hand in hand with open data, and so would many actions that could be taken to preserve its authenticity, unless they result in a reduction of the possibility to modify the data and mix it with other information. Integrity can be guaranteed with respect to open data infrastructures, but open data itself has to be in an interoperable, machine-readable, and open format – as modifiable and re-usable as possible, in good substance, which is a set of attributes that is not meant to favour integrity. Measures aiming at ensuring confidentiality by restricting the accessibility or findability of the information are strictly incompatible with open data by the latter’s very definition, and so would be restricting its acceptable uses only to certain purposes or endeavours. As far as confidentiality goes, it would seem that the only solutions that can be adopted without conflicting with open data principles are ex ante measures such as pseudonymisation and anonymisation. All in all, it appears that balancing open data and data protection would require forsaking the possibility to rely on many of the security measures through which personal data protection is implemented in practice.

There are, to be sure, some obligations to which data controllers are bound that would not necessarily need any form of dilution, if personal data protection were to be balanced with open data. Yet, many other would. In addition, as discussed, many of the data subject rights granted by the GDPR would be impossible to guarantee, or rendered practically moot, if personal data were shared according to open data’s requirements. Most importantly, when thinking about balancing open data and data protection, one should move from the fundamental incompatibility between the majority of the principles ex Art. 5 of the GDPR, which lists the core elements that characterised data protection laws worldwide for the past four decades, and open data’s main tenets. On the whole, the possibility to publish personal data as open data would require significant derogations, of either a statutory, judicial, or interpretative nature, to a large number of EU data protection rules and principles.

Could that happen, and should it happen? In other words, would a compression of the right to personal data protection, aiming at balancing it with the principles underlying open data, be suitable, necessary, and proportionate to the benefits it would achieve?

5.5.2 WOULD IT BE SUITABLE?

Both the Charter and the ECHR deal with lato sensu proportionality balancing by splitting it into the familiar suitability, necessity, and stricto sensu proportionality sub-tests, meant to ascertain whether the compression of a right – in this case, the compression of data protection by the rights and freedoms that can be upheld through open data sharing and re-use – can be deemed legitimate. Any limitation on the rights and freedoms recognised by the Charter must be provided for by law, and respect their essence. Rights and freedoms may be limited only if that is necessary, genuinely meets (i.e. is suitable for the achievement of) objectives of general interest recognised by the EU, or the need to protect the rights and freedoms of others, and is strictu senso proportionate to those objectives. Likewise, Art. 8(2) ECHR states, with a

1139 Charter, Art. 52(1).
1140 In this respect, Art. 52(3) of the Charter specifies that in so far as it contains rights that correspond to the ones guaranteed by the ECHR, the meaning and scope of the Charter’s rights is the same as those laid down by the ECHR.
formulation that is echoed in other articles that provide for the possibility to limit one of the Convention rights, 1141 that there shall be no interference by a public authority with the exercise of the right to private life, except where in accordance with the law, necessary in a democratic society, and suitable for the pursuit of a certain number of specific and yet widely framed objectives. 1142 In other words, a measure constituting an interference with a right is disproportionate by default if unsuitable for the achievement of the aims it pursues.

A first consideration is thus that, for an interference to be legitimate, it must be provided by law. 1143 Indeed, in the words of Advocate General Cruz-Villalón, “(i)n the absence of ‘law’ […] it is not in fact necessary to examine, in turn, the conditions to which any limitation on the exercise of the rights and freedoms recognised by the Charter is subject and in particular the condition of proportionality”. 1144 As Advocate General Saugmandsgaard Øe points out in his Opinion on the Tele2 Sverige case, until the 2015 WebMindLicenses judgement, 1145 CJEU case law did not explicitly provide a precise definition of the lawfulness requirement, rather relying on the requirements set by the ECtHR’s case law. 1146 In WebMindLicenses, the CJEU explicitly clarified, for the first time, that “the requirement implies that the legal basis […] must be sufficiently clear and precise and that, by defining itself the scope of the limitation on the exercise of the right guaranteed […] it affords a measure of legal protection against any arbitrary interferences”. 1147

Regardless, as mentioned, ECtHR case law has provided for a long time a list of the necessary attributes that the ‘provided for by law’ requirement must have, and EU jurisprudence has been able to rely on those requirements for the better part of half a century. A limitation is ‘provided for by law’, if it meets three cumulative conditions: “(f)irst, the limitation must have some basis in domestic law. […] Second, the persons concerned must be able to know those limitations in advance, […] Finally, those limitations must be foreseeable”. 1148 The ECtHR frequently held that the condition that an interference in the exercise of a right must be ‘provided for by law’ does not only mean that the measure must lay on a legal basis, 1149 but

---

4. In so far as this Charter recognises fundamental rights as they result from the constitutional traditions common to the Member States, those rights shall be interpreted in harmony with those traditions.

1141 E.g. Art. 10(2) ECHR. Under the ECHR, every qualified right (i.e. a right that may be subject to limitations, as opposed to absolute rights, which may not) is followed by a specific derogation clause. Under the Charter, Art. 52(1) functions as a general limitation clause, containing the conditions with which limitations on the exercise of every (qualified) right and freedom recognised by the Charter must comply: see Koen Lenaerts, ‘Exploring the Limits of the EU Charter of Fundamental Rights’ (2012) 8 European Constitutional Law Review 375, 388.

1142 I.e. national security, public safety, the economic well-being of the country, the prevention of disorder or crime, the protection of health or morals, or the protection of the rights and freedoms of others.

1143 The assessment of the legality of the measure is sometimes distinct from the assessment of its suitability; this thesis frames the legality test as the first step in the suitability assessment, but the two constructions are practically equivalent.

1144 Opinion of Advocate General Cruz-Villalón in Scarlet Extended SA v Société belge des auteurs compositeurs et éditeurs (SABAM), Case C-70/10, 14 April 2011 ECLI:EU:C:2011:255, § 37.


1146 Opinion of Advocate General Saugmandsgaard Øe in Tele2 Sverige AB v Post- och telestyrelsen (C-203/15) and Secretary of State for the Home Department v Tom Watson, Peter Brice, Geoffrey Lewis (C-698/15), Joined Cases C-203/15 and C-698/15, 19 July 2016, ECLI:EU:C:2016:572, §136.

1147 CJEU, WebMindLicenses (n 1145) §81.

1148 Lenaerts (n 1141) 389. See also Rainey, Wicks and Ovey (n 662) 343.

1149 See e.g. ECtHR, Sunday Times Case, 2 E.H.R.R. 245 (1979) §47; Malone v. United Kingdom (n 668) §66; Leander v. Sweden (n 851) §50; .
also that such basis must fulfil certain requirements relating to its quality. The law must, in particular, be accessible and foreseeable, precise enough to enable individuals to regulate their behaviour in light of what they can foresee, to a reasonable degree, as the consequences deriving from a given action or omission. The law must be clear, and its meaning and nature predictable; the scope and manner of the interference in the exercise of the rights guaranteed must be defined with sufficient precision. That does not mean that discretion is ruled out tout court, but rather that its scope and the manner of its exercise must be specified with sufficient clarity, in consideration of the aims it pursues, to provide sufficient protection against arbitrary interference. Limiting a right is hence possible only if the legal basis on which the limitation rests is accessible, clear, and foreseeable to the persons concerned, which depends, concretely, on the content of the limitation considered, the field it covers, and who it addresses. ECtHR jurisprudence holds that ‘provided for by law’ has a substantive meaning, rather than a merely formal one: it may include both written law and unwritten law, even case law, to some extent. The ECtHR’s interpretation of the meaning of ‘law’ is thus fairly extensive.

In relation to the suitability of a hypothetical measure aiming at balancing open data and data protection, the ‘provided for by law’ requirement implies the necessity of a legal basis. The matter is likely to fall within the shared competences between the EU and its Member States. The EU has already exercised its competence in the areas concerned, and the margin of discretion granted by the GDPR is not, as it is, broad enough to allow a restriction of the right to data protection sufficient to allow personal data to be shared and re-used as open data. It seems thus inevitable that the legal basis justifying such a restriction would have to

---

1151 See e.g. Sunday Times Case (n 1149) §47; Silver v. United Kingdom, 5 E.H.R.R. 347 (1983) §86; Leander v. Sweden (n 851) §50.
1152 See e.g. Sunday Times Case (n 1149) §49; Malone v. United Kingdom (n 668) §66; Leander v. Sweden (n 851) §50.
1154 See e.g. Sunday Times Case (n 1149) §49; Malone v. United Kingdom (n 668) §66; Silver v. United Kingdom (n 1151) §88; Leander v. Sweden (n 851) §51; Margareta and Roger Andersson v. Sweden (n 1150) §75.
1156 See e.g. Malone v. United Kingdom (n 668) §66; Silver v. United (n 1151) §88-89; Leander v. Sweden (n 851) §51.
1157 The CJEU seems to follow suit as, e.g. in the Eifert case (n 921, §66), it considers Council Regulation 252/2008 to fit the definition of law for the purposes of Art. 52(1) of the Charter, which means that it is not required for the legal basis that provides for a limitation of a fundamental right to follow the ordinary legislative procedure. In other words, ‘law’ is broad enough to encompass acts that are not conditioned to the Parliament’s codetermination: see Lenaerts (n 1141) 390, 391.
1158 Art. 2-6 TFEU.
1159 I.e. data protection on one hand, and PSI, access to documents, freedoms of business, and the various rights and freedoms that can be supported through open data sharing and re-use, on the other hand.
1160 See s. 5.1.1 above, particularly in light of the fact that well-established CJEU jurisprudence holds that exceptions and derogations to the right to personal data protection must be interpreted restrictively: see CJEU, Tele2 Sverige AB (C-203/15) v Post- och telestyrelsen, and Secretary of State for the Home Department (C 698/15) v Tom Watson, Peter Brice, Geoffrey Lewis, Joined Cases C-203/15 and C698/15, 21 December 2016, ECLI:EU:C:2016:970, §96, and the jurisprudence cited therein.
be an EU legal act – which does not however rule out Member States’ discretion entirely, if delegated by the act in question.

Aside from having to be provided for by law, limitations to fundamental rights must also be suitable for the achievement of certain objectives. The Charter states that limitations must “genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others”. The Charter’s explanations specify that “(t)he reference to general interests recognised by the Union covers both the objectives mentioned in Article 3 of the Treaty on European Union and other interests protected by specific provisions of the Treaties”. Art. 3 TEU sets out the EU’s fundamental values and objectives, which are complemented by several other provisions at a primary level; the explanations to the Charter mention explicitly only Art. 4(1) TEU and Art. 35(3), 36 and 346 TFEU, yet there are plenty of (quasi)constitutional provisions in the Treaties (and in the Charter, by incorporation) that enshrine EU goals and values, and the wording “objectives of general interest recognised by the Union” is thus to be seen as quite broad. The ECHR, differently, lists the objectives that may legitimize the compression of the right to private life explicitly: national security, public safety, the economic well-being of the country, the prevention of disorder or crime, the protection of health or morals, or the protection of the rights and freedoms of others.

There is little doubt that the open data movement, and open data as a data sharing policy, pursue objectives of general interests. The goals of transparency, accountability, efficiency, and technological and economic development are commonly associated with the open data ethos, and the benefits and value of open data and PSI have been postulated or proven throughout the past two decades. It may even be argued that open data sharing and re-use can further the pursuit and the defence of a wide array of fundamental rights and freedoms enshrined in both the ECHR and the Charter, such as freedom of expression and of

---

1161 Charter, Art. 52(1).
1162 Charter, Explanation on Article 52 — Scope and interpretation of rights and principles.
1163 The explanations themselves do not have the status of law, and yet they are a valuable tool to interpret the Charter’s provisions - Lenaerts (n 1141) 401. Moreover, the explanations have been explicitly referred to by the CJEU’s case law: see e.g. Case C-279/09, DEB Deutsche Energiehandels- und Beratungsgesellschaft mbH v Bundesrepublik Deutschland. 22 December 2010, ECLI:EU:C:2010:811, §32, 39.
1164 The choice of the articles exemplified would appear to be owed to the necessity to re-assure individual Member States about the fact that the Charter is not meant to lower the degree of national autonomy each Member State has.
1165 Lenaerts (n 1141) 391.
1166 ECHR, Art. 8(2).
1167 Cfr. CJEU, European Commission v The Bavarian Lager Co. Ltd., Case C-28/08 P, 29 June 2010, ECLI:EU:C:2010:378, §54: “openness enables citizens to participate more closely in the decision-making process and guarantees that the administration enjoys greater legitimacy and is more effective and more accountable to the citizen in a democratic system”.
1168 See Zuiderveen Borgesius, Van Eechoud and Gray (n 41); Scassa (n 614).
1170 E.g. Vickery (n 404); Henninger (n 29); De Vries and others (n 404); Carrara and others (n 31).
information, access to documents, and freedom of business. From this perspective, it is undeniable that the aims pursued through open data sharing and re-use are in line with the general interests recognised by the EU, or with the “economic well-being of the country” – albeit potentially to the detriment of other competing rights and freedoms.

Is open data also suitable for their achievement?

Intuitively, one would think that open data is, ideally, suitable to achieve – or, to the very least, to pursue – all the objective of general interest that can be attained through less permissive sharing options and use policies. In other words, prima facie, it could seem that all the goals that can be reached through an instrument mandating a form of mediated or constrained data sharing and use can also be achieved through direct, unconstrained access to bulk data in a machine-readable format. That seems indeed to be the case, particularly in the age of digital start-ups, citizen science, and data journalism, where information literacy has been on the steady rise. Yet, depending on how such an instrument is framed – i.e. depending on the construction of the legal basis that would justify a compression of the right to personal data protection that would allow personal data to be shared as open data – some doubts may still be cast.

The latest instalment of the Satamedia saga provides for a concrete example of how the suitability of open data on its own, and the release of unaltered bulk information in which it substantiates, could potentially be called into question. One of the main points of the case revolved around if and to which extent the bulk publication of raw taxation data pertaining to a multiplicity of individuals can fall under the derogations for journalistic activities granted by Finnish data protection law. In this respect, the Finnish government argued that the publication of bulk, raw data, without any accompanying analysis, cannot be viewed as ‘data journalism’, due to the lack of conclusions deriving from the analysis of the data which could have contributed to a debate of public interest. By analogy, one might argue that open data’s suitability to achieve its high-level goals (transparency, accountability, efficiency, growth) does not derive only from the availability of the information in itself, but also from e.g. its controls or use restrictions might also impact the freedom of business of companies already accessing and using such information.

1171 Charter, Art. 11.
1172 Charter, Art. 41.
1173 Charter, Art. 42.
1174 Charter, Art. 16.
1175 Charter, Art. 52(1).
1176 Art. 8(2) ECHR.
1177 Privacy and data protection, for instance; a measure meant to open up a dataset previously subject to access controls or use restrictions might also impact the freedom of business of companies already accessing and using such information.
1178 Cfr. CJEU, Michael Schwarz v Stadt Bochum, Case C 291/12, 17 October 2013, ECLI:EU:C:2013:670, §43, where the court held that appropriateness/suitability does not require perfect results: “it must be held that the fact that the method is not wholly reliable is not decisive”.
1179 ECtHR, Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland (n 300).
1180 The affaire, involving Satakunnan Markkinapörssi Oy and Satamedia Oy, is a series of cases about the competition between Art. 8 and Art. 10 of the ECHR (and hence between Art. 7 and 8 of the Charter, on the one hand, and Art. 11 on the other) and the scope of the journalistic exception in EU data protection law, which spanned from the early 2000s to 2017, and went from Finnish administrative courts to the CJEU (CJEU, Tietosuojavaltuutettu v Satakunnan Markkinapörssi Oy, Satamedia Oy (n 1122)) and eventually to both the ECtHR’s Chamber and its Grand Chamber (ECtHR, Tietosuojavaltuutettu v Satakunnan Markkinapörssi Oy and Satamedia Oy, 2008 E.C.R. I. 9831 (2008); ECtHR, Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland (n 300)).
1181 ECtHR, Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland (n 300) §110.
governance or from the user’s characteristics. From this perspective, the suitability of a measure mandating or allowing the bulk publication of personal data to achieve the objectives of general interest (or the protection of the rights and freedoms of others) underlying it might not be taken for granted, and be made to depend on other contextual elements. In other words, there may be room to argue about the suitability of the publication of data for its own sake for the achievement of the objectives of general interest that must justify the compression of the right to data protection that would happen if personal data were to be shared as open data. That would however be, to a large extent, playing the devil’s advocate. The open data ethos revolves around objectives of general interest, and open data is certainly as suitable to their achievement as more restrictive information sharing and re-use options are. Provided that the instrument constituting the legal basis for a compression of the right to personal data protection satisfies the quality requirements, it seems difficult to argue about open data’s suitability to reach the policy goals set. That appears to be the case particularly in light of the fact that courts, generally, when performing the suitability sub-test within their lato sensu proportionality assessment, focus mostly on the (abstract) suitability of the measure’s purposes, rather than on their results.

Issues however arise when considering the idea of balancing open data and data protection through the lenses of the remaining requirements of the balancing test: necessity and stricto sensu proportionality. As the following sections argue, releasing personal data as open data is not really necessary, as data protection law does not really impede personal data processing, and there are more nuanced data sharing and re-use options that can achieve the same goals within the boundaries of current EU data protection legislation. Furthermore, even if the smart city narrative were successful in advancing the argument that the legitimisation of some edge cases of personal data processing that do not meet the standards set by EU data protection law are “necessary” for the pursuit of an objective of general interest, publishing and allowing the re-use of such personal information as open data would not be proportionate by definition.

5.5.3 IS IT NECESSARY?

An interference upon a right, such as the one bound to happen if personal data were to be shared as open data, must be necessary. According to the Charter, “limitations may be made only if they are necessary and genuinely meet objectives of general interest”, while the ECHR states, with respect to the right to private life, that “(t)here shall be no interference […] except such as is […] necessary in a democratic society”. After the suitability sub-test, where the measure compressing a right is assessed in light of the objectives it aims at achieving, both courts and the legislature (should ideally) evaluate whether such measure is genuinely necessary in a democratic society. “Necessity implies the need for a combined, fact-based assessment of the effectiveness of the measure for the objective pursued and of whether

\[1182\] Welle Donker and van Loenen, on a similar note, contest the suitability of current open data assessment methodologies, and in their research, use three output indicators as conditions for a successful open data ecosystem: data supply (the way in which data are provided as open data); data governance (the way in which governance aspects are organised); and user characteristics (the way in which the user characteristics enable the user to innovate with open data) – see Welle Donker and van Loenen (n 365).

\[1183\] With regard to the ECtHR, see Christoffersen and Wahlgren (n 1011) 31.

\[1184\] Charter, Art. 52(1).

\[1185\] Art. 8(2) ECHR.
it is less intrusive compared to other options for achieving the same goal”.

Both suitability and necessity thus move from an assessment of the effectiveness of the measure in respect to the objectives it aims at achieving. While the suitability test evaluates the logical link between purpose and means, however, the necessity test warrants a factual description of the measure concerned, intending to ascertain whether there are less-interfering options available that could still achieve the purposes set.

5.5.3.1 The factual description of the interference

In respect to the content that the factual description of a measure limiting the right to personal data protection should have, the EDPS Necessity Toolkit provides a convenient and synthetic (albeit non-exhaustive) overview. The toolkit identifies, amongst the elements that should make up the measure’s factual description, its objective and the general interest underlying them; the exact purpose of the processing, who is supposed to process the data, and which processing operations are foreseen, amongst any other relevant element, such as the expected duration of the processing. It seems evident, when comparing those elements with the requirements that information must have to qualify as open data, that the factual description of the measure requires a level of specificity in setting the boundaries of the processing that would be difficult to attain. Open data requires data to be usable by anyone, for any purpose, without access controls or use restrictions. The purposes of the entity distributing the information, for instance, would have to be framed with more precision than ‘publishing information for publishing’s sake’, and the anticipated purposes for re-users would still (have to) encompass any purpose that the data in question could possibly have.

The EDPS Necessity Toolkit also includes, amongst the elements that a measure interfering with data protection should have, the categories of personal data and the types of data subject that would be involved by such a measure. That appears both compatible with open data requirements and an essential condition for the idea of balancing open data and data protection to be even taken into consideration. Considering the quality requirements (clarity, foreseeability, accessibility) that a measure restricting a right must have, in conjunction with the complex and wide material scope of EU data protection law and the severity of the limitations to data protection that releasing personal information as open data would entail, it seems inevitable for it to indicate with precision the specific kind or categories of personal data interested by such a measure, and the categories of controllers involved. By virtue of the extent to which releasing personal data as open data compresses data subjects’ right to data

---


1187 Since both suitability and necessity are part of the broader lato sensu proportionality assessment, both are tested against the objectives of general interest (or the protection of the rights and freedoms of others) pursued, and the structure of proportionality balancing is far from set in stone, there are inevitable overlaps between the content of proportionality’s three sub-tests.

1188 European Data Protection Supervisor (n 1186).

1189 “(E)xplained in more detail than the objective” - European Data Protection Supervisor (n 1186) 10.

1190 See Open Knowledge International (n 32).

1191 European Data Protection Supervisor (n 1186).

1192 See section 6.2 above.

1193 See e.g. Brkan, ‘The Unstoppable Expansion of the EU Fundamental Right to Data Protection: Little Shop of Horrors?’ (n 903); van Loenen, Kulk and Ploeger (n 977); Purtova, ‘The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law’ (n 903).

1194 See section 6.1 above.
protection, and of the number of data protection rights and principles that would need to be limited, it seems that a broader, “horizontal” derogation\textsuperscript{1195} would have, fundamentally, the effect of deregulating personal data processing in general.\textsuperscript{1196}

5.5.3.2 A PERMISSIVE REGIME

The previous chapter, in discussing this thesis’ conception of the substance and essence of data protection, subscribed to its framing as a permissive\textsuperscript{1197} tool of transparency,\textsuperscript{1198} as opposed to a prohibitive tool of opacity. The nature of data protection, and the fact that it is meant to permit (and yet channel) personal data processing, rather than curtail it, is both highlighted by and resulting from a plethora of diverse provisions within the GDPR. Indeed, at its very outset, the GDPR clarifies how its subject matter and objective\textsuperscript{1199} – in line with its main antecedents, Convention 108 and the DPD – is not only the protection of individuals with regard to the processing of their personal data, but also the free movement of such data. The GDPR protects, broadly, people’s fundamental rights and freedoms, in particular their right to the protection of personal data. It also mandates, however, that the free movement of personal data within the EU must not be either restricted or prohibited by virtue of personal data protection. Modern EU data protection law’s raison d’être is thus split between protection and permission, and underlines the constant tension between its fundamental rights dimension and its market-making rationale.\textsuperscript{1200}

The permissive nature of data protection results with clarity from the analysis of the grounds that make personal data processing legal, in compliance with the lawfulness principle.\textsuperscript{1201} Processing is lawful only if and to the extent that it is legitimised by (at least) one of the following grounds: the data subject’s consent, the performance of a contract to which the data subject is party,\textsuperscript{1202} compliance with a legal obligation to which the controller is subject, the vital interests of the data subject or of another natural person, the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller, and the legitimate interests of the controller or of a third party.\textsuperscript{1203} Amongst those grounds, consent is the only one that requires the exercise of the data subject’s own decisional power: regardless of the undeniable link between data protection and informational self-determination,\textsuperscript{1204} EU data protection law is modelled along the lines of Convention 108, which goes well beyond consent.\textsuperscript{1205}

\begin{enumerate}
\item \textsuperscript{1195} Cfr. the characteristics of data protection law as an omnibus regime, as described in Lynskey, \textit{The Foundations of EU Data Protection Law} (n 19).
\item \textsuperscript{1196} See, by analogy and \textit{mutatis mutandis}, CJEU, Tietosuojavaltuutettu v Satakunnan Markkinapörssi Oy, Satamedia Oy (n 1122) §48: “it must be held that a general derogation from the application of the directive in respect of published information would largely deprive the directive of its effect. It would be sufficient for the Member States to publish data in order for those data to cease to enjoy the protection afforded by the directive”.
\item \textsuperscript{1197} See Fuster and Gutwirth (n 882); Lynskey, \textit{The Foundations of EU Data Protection Law} (n 19) 30.
\item \textsuperscript{1198} See De Hert and Gutwirth (n 858); De Hert and Gutwirth (n 236).
\item \textsuperscript{1199} GDPR, Art. 1.
\item \textsuperscript{1200} See Janciute (n 787); Lynskey, \textit{The Foundations of EU Data Protection Law} (n 19) 46 ss.
\item \textsuperscript{1201} GDPR, Art. 5(1)(a).
\item \textsuperscript{1202} Or, if requested by the data subject, taking steps prior to entering into a contract.
\item \textsuperscript{1203} Except where the controller’s interests are overridden by the interests, rights, or freedoms of the data subject, in particular where he or she is a child.
\item \textsuperscript{1205} And thus way beyond informational self-determination: see Hustinx (n 663).
\end{enumerate}
There are thus several grounds that can be used to justify the publication and the (re)use of personal data without having to ask data subjects for their consent. Public-sector bodies can resort to the grounds given by the necessity to comply with a legal obligation to which they are subject, or the performance of a task carried out in the public interest or in the exercise of the official authority they vest, provided that there is an appropriate legal basis in either EU or Member State law. Public-sector bodies cannot rely on the legitimate interest ground to justify personal data processing carried out in the performance of their tasks, but re-users and other publishers of information that do not belong to the public sector can. On the condition that the result of the balancing test that must be performed when relying on legitimate interest favours the controller’s interests over the data subject’s, there’s nothing standing between re-users and compliance with the lawfulness principle. There are plenty of ways for both data publishers and (re)users to comply with the lawfulness principle when sharing and re-using personal data, if convenient – let alone necessary – for whichever objective smart city stakeholders might have: the choice is not necessarily between open and closed data, but rather between different shades of open.

Sensitive data – what the GDPR calls “special categories of data” – would seem to be subject to a different, prohibitive regime: Art. 9(1) mandates that the “(p)rocessing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person’s sex life or sexual orientation shall be prohibited”. Art. 9(2) however provides for a very wide set of exceptions to the prohibitive regime hinted at by the preceding paragraph, many of which are applicable to smart city objectives or to enable personal data processing to


1207 GDPR, Art. 6(1)(c) and 6(1)(e).

1208 GDPR, Art. 6(3). A fitting example could be the Dutch national law regulating the tasks and obligations of the cadastre vis-à-vis geo-information publishing, about which I refer to Anna Berlee, Access to Personal Data in Public Land Registers: Balancing Publicity of Property Rights with the Rights to Privacy and Data Protection (Eleven international publishing 2018).

1209 GDPR, Art. 6(1). It is questionable whether public sector bodies might be able to do so when acting outside of the boundaries of the performance of their tasks.


1211 Cfr. CJEU, Camera di Commercio, Industria, Artigianato e Agricoltura di Lecce v Salvatore Manni, Case C-398/15, 9 March 2017, §42-43: “it should be noted that the processing of personal data by the authority responsible for keeping the register […] satisfies several grounds for legitimation provided for in Article 7 of Directive 95/46 […] it should be noted that the Court of Justice has already held that the activity of a public authority consisting in the storing, in a database, of data which undertakings are obliged to report on the basis of statutory obligations, permitting interested persons to search for that data and providing them with print-outs thereof, falls within the exercise of public powers […]. Moreover, such an activity also constitutes a task carried out in the public interest within the meaning of that provision”.

pursue the goals underlying open data sharing. Consent aside, and regardless of the fact that Art. 9(1) does not apply to the processing of personal data which are manifestly made public by the data subject, processing of sensitive data can be justified if necessary for reasons of substantial public interest, on the basis of EU or Member State law, or for archiving, scientific, historical research, or statistical purposes. All in all, the difference between the default regime and the one applicable to sensitive data is thinner than the wording of Art. 9(1) would suggest.

The purpose limitation principle has traditionally been portrayed as being at odds with open data. The idea that publishing information for its own sake, under the Silicon Valley-sh belief that “if you publish it, they will come”, is inherently in conflict with the principle for which personal data must be “collected for specified, explicit and legitimate purposes”. That may very well be the case, but it is still for data controllers to decide the purpose of their processing. Granted, it may be difficult (which, incidentally, does not mean ‘impossible’) for data publishers to articulate the purpose of the processing of ‘serendipitous’ data sharing. Yet, if they (can) resort to the grounds for lawful processing given by the necessity to comply with a legal obligation, or to perform a task carried out in the public interest or in the exercise of the controller’s authority, then the purpose of the processing is determined by the legal basis of the obligation, or by the nature and normative underpinning on the task carried out, so specifying the processing’s purpose stops being an issue. If the processing is not based on Art. 6(1)(c) or 6(1)(e), again, data controllers still gets to decide its purpose: if they cannot do so in a satisfactory manner, perhaps they should not deal with personal data in the first place.

Furthermore, the purpose limitation principle does not really mean that personal data cannot be processed for purposes different from the original collection’s, but just that they must not be “further processed in a manner that is incompatible with those purposes”. The prohibition is represented by a sort of double negation: further processing does not need to be compatible with the collection’s purposes, but just not incompatible, which is quite a different threshold. Notably, as it fits well with the goals underlying the open data movement, further processing for archiving purposes in the public interest, scientific or historical research purposes, or statistical purposes are not to be considered incompatible with the initial purposes. Further processing can be based on consent, or on EU or Member State law, in

1213 Which must be proportionate to the aims pursued, respect the essence of the right, and provide for suitable and specific measures to safeguard the fundamental rights and the interests of data subjects: GDPR, Art. 9(2)(g).
1214 In accordance with Article 89(1), and based on an EU or Member State law that is proportionate, respects the essence of the right to personal data protection, and provides for suitable and specific measures to safeguard the fundamental rights and the interests of data subjects.
1215 And with big data in general – see Zarsky (n 1094).
1216 “If you build it, they will come”, a play on a quote (“if you build it, he will come”) from the 1989 movie “Field of Dreams”, is one of the most notorious ‘memes’ used to describe the ‘startup milieu’ and the related Silicon Valley ethos.
1218 Further processing of personal data for research purposes, for instance, enjoys a legal presumption of non-incompatibility in the GDPR: see Art. 5(1)(b).
1219 GDPR, Art. 6(1)(c).
1220 GDPR, Art. 6(1)(e).
1222 GDPR, Art. 5(1)(b), last sentence.
which case the compatibility (rectius: non-incompatibility) of the further processing instance is given by the data subject’s volition or by the legal basis legitimising it. Otherwise, the controller can determine compatibility with the original purpose on its own, taking into account (amongst other things) the link between the original purposes and the purposes of the further processing, the context in which the personal data have been collected, the nature of the personal data, the possible consequences for data subjects, and the existence of appropriate safeguards. All in all, purpose specification and limitation do not really impede the kind of personal data processing that is necessary for cities to run, regardless of whether smart or not, nor curtail the rights, freedoms, and interests underlying open data, but rather balances them with the competing right to personal data protection.

The ties between the purpose specification principle and other principles such as data minimization and storage limitation make it so that the ability to decide the processing’s purposes, or to rely on purposes set by EU or Member State law, influences how those other principles play out in concrete cases. The obligation to comply with data minimization and storage limitation can be tuned through the purposes set by the controller or by the EU or Member State law that provides for the legal basis of the processing. The kind of ‘purposeless’ processing required by a strict definition of open data impacts the data minimization and storage limitation principles as well, but the opposite also holds true: data controllers (and legislatures) can, to a certain extent, determine the minimization threshold and the storage limit by adjusting the processing’s purposes. In this sense, the function of the purpose specification, data minimization, and storage limitation principles has more to do with enhancing accountability, transparency, and legal certainty in personal data processing, than with curtailing it.

There is no denying that compliance with EU data protection law on the ground can be a high-intensity effort, and that data controllers are subject to a multiplicity of obligations that require the expenditure of time and resources to be respected. It is however also true that plenty of obligations (and, on the other side of the medal, of data subject rights) present exemptions and derogations that modulate many of the obligations arising from the GDPR in light of the characteristics of the kind of processing they regulate, and of the risks it presents. Aside from the specific exemptions and derogations contained in Art. 12-22 on data subject rights and in Art. 24-36 on data controller obligations, there are two set of norms that highlight in a particular way how the GDPR – and the DPD before it – is structurally primed to accommodate categories of processing that are necessary for the pursuit of objectives of public interest: Art. 23 and Art. 85-91.

Art. 23 allows EU or Member State law to restrict, through a legislative measure, the scope of the data subject rights (and the countervailing controller obligations) ex Articles 12-22. Such restriction can compress the data subjects’ rights to the point of limiting the core principles set

---

1223 GDPR, Art. 6(4).
1224 See above s. 5.5.1.2.
1225 On the difference between privacy on the books and privacy on the ground, see Kenneth A Bamberger and Deirdre K Mulligan, Privacy on the Ground: Driving Corporate Behavior in the United States and Europe (MIT Press 2015).
1226 See generally Article 29 Data Protection Working Party, ‘Statement 14/EN WP 218 on the Role of a Risk-Based Approach in Data Protection Legal Frameworks’ (2014). The so-called ‘risk-based approach’ has often been touted as a novelty introduced from the EU data protection law reform process started in 2012; yet, naming aside, it has (more) convincingly been argued that
1227 And the communication of a personal data breach to the data subject ex Art. 34.
by Art. 5, but only to the extent that it corresponds to the data subject rights and data controller obligations provided for in Art. 12-22. The legitimacy of such restriction is conditional to the respect of the essence of the fundamental rights and freedoms concerned, to the necessity and proportionality of the measure involved, and to the pursuit of one or more of the objectives listed in Art. 23. Those objectives are enumerated exhaustively, but are also framed quite broadly, and the goals and aims that underlie the concept of smart city can surely be portrayed as pursuing them. Yet, measures restricting data subject rights must have clear boundaries, and the provisions in which they are contained must specify, where relevant and amongst other things, the purposes of the processing, the type(s) of personal data, the scope of the restrictions and the relative safeguards, the specification of the controller or categories of controllers, the storage periods, and the applicable safeguards in light of the risks to the rights and freedoms of data subjects. The GDPR, in other words, already foresees the possibility to limit the rights and freedoms of data subjects if necessary for (and proportionate to) a multiplicity of objectives that (can be framed as to) include the ones on which the smart city narrative is based. It does so, however, by limiting the provisions that can be object of derogation, and by considering the provisions it must contain to edge the extent of such derogation.

Similarly, Art. 85-91 contain provisions regulating specific processing situations that are deemed to be particular, in that they present high social value or are peculiar enough to necessitate a special regime. Those provisions cover a wide range of situations, going from the processing of personal data in the context of employment relationships to the rules applicable to churches and other religious organisations. For the purposes of this thesis, it seems important to highlight the presence of three distinct articles, and the absence of one provision. Art. 85 of the GDPR covers personal data processing instances that involve the often-competing freedom of expression and information, stating that Member States must, by law, adopt measures aiming at the reconciliation of personal data protection with freedom of expression and information, including – and particularly – for journalistic, academic, artistic or literary purposes. For processing carried out for those purposes, Member States can indeed provide for exemptions or derogations from the principles, rights, obligations, and general mechanics contained in the GDPR, if it is necessary to square data protection and freedom of expression/information. Likewise, Art. 86 of the covers personal data processing vis-à-vis the right to access. It allows, in accordance with EU or Member State law, the disclosure of

---

1228 E.g. under “other important objectives of general public interest of the Union or of a Member State, in particular an important economic or financial interest” (Art. 23(1)(e)), which also enables restrictions aiming at “monitoring, inspection or regulatory function connected, even occasionally, to the exercise of official authority” connected with those interests (Art. 23(1)(h)).
1229 GDPR, Art. 23(2).
1230 GDPR, Art. 88.
1231 GDPR, Art. 91.
personal data contained in official documents held by a public authority, a public body, or a private body for the performance of a task carried out in the public interest, in order to reconcile public access to documents with the right to data protection.

The possibility to derogate the standard GDPR regime is also foreseen for activities relating to archiving purposes in the public interest, and scientific, historical, or statistical purposes. By virtue of Art. 89, for scientific, historical research, or statistical purposes, EU or Member State law may derogate from the rights of access, rectification, restriction, and objection. Those derogations may be adopted only if essential for the fulfilment of those purposes, and in so far as such rights may render impossible, or otherwise seriously impair, the achievement of the scientific, historical research, or statistical purposes.\footnote{GDPR, Art. 89(2).} For archiving purposes in the public interest, EU or Member State law may derogate from the right of access, rectification, restriction, and objection (in addition to the notification \textit{ex} Art. 19), and also from the right to personal data portability, in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.\footnote{GDPR, Art. 89(3).} In any case, specific processing situations relating to archiving purposes in the public interest, and scientific, historical, or statistical purposes must be made subject to appropriate technical and organisational safeguards meant to protect data subjects’ rights and freedoms, especially with regard to the data minimisation principle.\footnote{GDPR, Art. 89(1).} Notably, where those kinds of processing have, at the same time, a different purpose, the derogations apply only to processing for archiving purposes in the public interest, and scientific, historical, or statistical purposes.\footnote{GDPR, Art. 89(4).}

It is noticeable how the derogatory regime foreseen for special processing situations by Art. 85, 86 and 89 of the GDPR includes many of the goals and objectives that are generally portrayed as advanced by the open data movement – e.g. transparency, accountability, public-sector efficiency, and the facilitation of knowledge discovery and production. With the exception of the ‘innovation and growth’ angle, where open data is used in support of freedom of business, and which falls under the conceptual scope of EU PSI legislation. The absence of a derogatory system concerning PSI re-use in the GDPR (an option that was considered,\footnote{See Jaatinen (n 1061).} but not adopted, during the EU data protection reform\footnote{De Hert and Papakonstantinou (n 802); De Hert and Papakonstantinou (n 807).} process), in conjunction with the provisions of the PSI Directive that provide for an exemption from the general PSI regime (i.e. mandate the opposite of what a PSI exemption in the GDPR would dictate), is indicative of how the legislature sees the interaction between freedom of business and the fundamental right to personal data protection.

The balance between competing rights is tuned, in secondary legislation, according to the function of each right in a democratic society. Freedom of expression and information, the right to access, and the right to good administration – many of the rights and freedoms underlying open data – are weighted against the rights and freedoms of the data subject in a different way than freedom of business. The reason is not hierarchy: fundamental rights are meant to be on an equal footing (although an argument could be made about the higher
standing of absolute rights vis-à-vis qualified rights). Rather, the reason may be sought in the original contrast between the fundamental rights dimension and the economic, market-making underpinning of EU data protection: in other words, EU data protection secondary legislation is the balance between freedom of business and personal data protection, and the norms containing exemptions and derogations to its general regime are meant to balance data protection with other competing rights and freedoms, such as freedom of expression and information or access to public documents.

To sum things up, the right to personal data protection, as it is currently substantiated through secondary legislation, is already a permissive right. The conditions that make personal data processing lawful are manifold and varied, and in particular the possibility to rely on the legitimate interest of the data controller or on a legislative measure that confers it a public task or a legal obligation opens up the legitimacy spectrum to all the kinds of processing a city needs to run, be it a smart one or not. Furthermore, despite having the legal efficacy (and thus the harmonising power) of a Regulation, the GDPR still contains a plethora of exceptions, exemptions, and derogations, allowing EU or MS law to enact, if necessary, a partially different regime. All in all, data protection is meant to allow personal data sharing, use, and re-use, provided that the conditions in which it substantiates in secondary legislation are met. Most of those conditions constitute a system of checks and balances that, by channelling processing through a set of rules and principles, balance the data subjects’ rights and freedoms with the individual and collective benefits that (can) derive from personal data processing. Checks and balances are fundamentally incompatible with the concept of open data, as they reduce the possibility to access and re-use information; they are not however incompatible with the goals and objectives pursued by the open data movement, nor with the rights and freedoms underlying them.

5.5.3.3 A BLATANT LACK OF NECESSITY

The outcome of the necessity test depends, concretely, on which threshold is set by the relevant jurisprudence. When it comes to limitations of the right to personal data protection, CJEU jurisprudence generally applies a strict necessity test; the ECtHR’s case law on the right to private life is less consistent, but the court still applies a strict necessity check in many cases revolving around personal data processing. The (partial) distinction might be due to the difference between the nature of the Strasbourg and the Luxemburg court: while the former is influenced by the minimum nature of the ECHR, and by the need to respect the margin of discretion individual member states have, the CJEU is concerned with ensuring a uniform interpretation and application of EU law.

Yet, regardless of the threshold adopted – be it the least restrictive measure (LRM) test, manifest unnecessary, or mere common sense – it is hard to argue about the necessity of balancing open data per se and data protection. Open data is not a right or freedom – it is a

---

1240 See Janciute (n 787); Lynskey, ‘From Market-Making Tool to Fundamental Right: The Role of the Court of Justice in Data Protection’s Identity Crisis’ (n 786); Lynskey, The Foundations of EU Data Protection Law (n 19) ss.

1241 The necessity test is also performed to draft and assess legislative proposals; the criteria used for this kind of lato sensu proportionality (thus including necessity) assessment are however still derived from the relevant courts’ jurisprudence.

1242 See e.g. CJEU, Tietosuojavaltuutettu v Satakunnan Markkinapörrä Oy, Satamedia Oy (n 1122); CJEU, Volker und Markus Schecke and Eifert (n 921); Tele2 Sverige (n 1160) §107 ss. See also Tranberg (n 1036).

1243 Christoffersen and Wahlgren (n 1011) 23 ss.
data sharing and re-use policy that may be adopted to advance underlying rights and freedoms. Data protection is a qualified right, not an absolute one; it has to be balanced with opposing rights and freedoms, including (and, perhaps, particularly) the ones on which open data is based. That appears particularly evident when looking at how the right to the protection of personal data is substantiated at a secondary level: the GDPR’s whole point is the pursuit of the fine balance between permitting the free flow of personal information and ensuring a high level of protection to the individuals to whom such information refers.

Assessing the necessity of balancing open data and data protection under the LRM test hardly requires much thought. Open data is at the most extreme end of the openness spectrum and thus every other less-permissive data sharing option constitutes a lesser compression of the right to personal data protection. That holds true if one considers the mere processing of personal data to be a compression of the right to data protection, as the CJEU sometimes does, but also if one frames an interference with the right as happening only if the processing deviates from data protection rules and principles, as this thesis does, since open data rejects the limits to access and (re)use in which those rules often substantiate. Put it differently, whichever data sharing and re-use option that foresee any degree of access control or use restriction, and is thus incompatible with the core tenets of open data, is a less restrictive measure.

I contend, moreover, that balancing open data and data protection is manifestly unnecessary for the achievement of whichever (legitimate) objective a ‘smart city’ could set: let alone the LRM test, the idea would not meet any sort of necessity threshold. The pursuit of the goals of the open data movement does not compel the processing of personal data, and the quest for the smart city, while arguably requiring the processing of personal data, does not really necessitate the publication of such data as open data. It may be argued that cities cannot be really ‘smart’ without the processing of personal data, which seems intuitively the case, but EU data protection law does not impede personal data processing for purposes connected to urban governance, and it is thus not necessary to amend it.

---

1244 See Bonnici (n 935); Gloria González Fuster, ‘Curtailing a Right in Flux: Restrictions of the Right to Personal Data Protection’ in Artemi Rallo Lombarte and Rosario García Mahamut (eds), *Hacia un Nuevo derecho europeo de protección de datos, Towards a new European Data Protection Regime* (Tirant Lo Blanch 2015).

1245 Cfr. (albeit with reference to necessity under Art. 7(1)(f) of the DPD rather than to the necessity of a measure compressing a right) CJEU, Heinz Huber v Bundesrepublik Deutschland (n 1118) §50, where the Court – after confirming the legitimacy of the statistical purposes pursued through the processing of immigration data on a centralised register, also notes that “the exercise of that power does not, of itself, mean that the collection and storage of individualised personal information in a register [...] is necessary, within the meaning of Article 7(e) of Directive 95/46 [...] it is only anonymous information that requires to be processed in order for such an objective to be attained”. The CJEU insists particularly in underlining that point (§68): “The storage and processing of personal data containing individualised personal information in a register [...] for statistical purposes cannot, on any basis, be considered to be necessary within the meaning of Article 7(e) of Directive 95/46”.

1246 E.g. in CJEU, Volker und Markus Schecke and Eifert (n 921).

1247 See Fuster and Gutwirth (n 882); Fuster and Gellert (n 882).

1248 An example of LRM testing (with a negative outcome) is given by CJEU, Volker und Markus Schecke and Eifert (n 921) §83, where the Court noted that the institutions ought to have examined, when balancing the interests involved, whether a more limited way of publishing information would have been sufficient to achieve the objectives set by the legislation at issue. The Court considered that “it does not appear that such a limitation, which would protect some of the beneficiaries concerned from interference with their private lives, would not [...] achieve the objectives of that legislation”. For a LRM test that did meet the necessity threshold, see CJEU, Michael Schwarz v Stadt Bochum (n 1178).
It may also be argued that cities cannot be really ‘smart’ without open data, which seems a fair argument, but then again, the vast majority of such data is not personal in nature. However, arguing for the necessity of allowing open personal data, lest we forego the cities of the future, appears to be quite a stretch. Personal data is a relatively small subcategory amongst all the information that can potentially be made open for smart city purposes, and the fact that it cannot be shared and re-used as open data, as data protection norms are incompatible with open data requirements, does not mean that it cannot be shared and re-used at all – quite the opposite.

There are three orders of reasons for which one might argue in favour of ‘balancing’ open data and data protection by resorting to the ‘smart city’ narrative, and none of them has to do with its strict necessity in a democratic society. The first one is the idea that deregulating personal data processing would lead to a positive trade-off, and that the negatives of EU data protection law outweigh its positives, which is not a legal argument, but a policy one, and has nothing to do with balancing and proportionality in the normative sense. The second reason is that compliance with data protection law can be difficult or burdensome for data controllers, particularly for smaller organisations, who often do not have in-house expertise on data protection matters. One might be sympathetic with such arguments – one might not – but, regardless of their validity, mere convenience is not strict necessity.

The third order of reasons, differently, does have a solid underpinning, and has to do with legitimate concerns about the over-inflation of the material scope of personal data. If identifying people is becoming easier and easier, anonymising data is becoming harder and harder, and inferring information about a person from seemingly unrelated data is easier and easier, then categories of information that have so far been well outside the scope of EU data protection law (e.g. climate data, or maps) could become personal data, with all the legal consequences that come with it.

If that were the case, many categories of information that are currently ‘safe’ for release as open data could be considered personal, which would have unwarranted side effects on the open data movement, and – if one subscribes to the idea that open data is a core element of smart cities – on the smart city’s development as well. Yet again, that has nothing to do with balancing and with lato sensu proportionality: it is rather about the interpretation of the

1249 See chapter 2, section 4.
1250 And arguably a bad one at that, as it goes against the grain of the course of action suggested by the growing (soft, media, and market) power of technology companies – see Orla Lynskey, ‘Grappling with “Data Power”: Normative Nudges from Data Protection and Privacy’ (2019) 20 Theoretical Inquiries in Law 189; Shoshana Zuboff, ‘Big Other: Surveillance Capitalism and the Prospects of an Information Civilization’ (2015) 30 Journal of Information Technology. See also Christl and Spiekermann (n 254); Christl, Kopp and Riechert (n 254). On the rise of surveillance capitalism, which would suggest an approach diametrically divergent from deregulating personal data processing, see generally Zuboff, The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power (n 98).
1252 See Rainey, Wicks and Ovey (n 662) 359 ss.
1253 Ohm (n 42).
1255 As argued by Purtova, ‘The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law’ (n 903).
1256 See van Loenen, Kulk and Ploeger (n 977); Kulk and Van Loenen (n 39).
concept of personal data, and about the translation of the state of the art of technical disciplines (e.g. statistical disclosure control, and computer science) into law and policy. Regardless, the inflation of the concept of personal data is a legitimate concern, and will be discussed extensively in the following chapter.

5.5.4 Could it be proportionate?

The comparatively lower amount of jurisprudential guidance on proportionality stricto sensu, as opposed to the other elements of the lato sensu proportionality test, is a fairly common complaint amongst academic lawyers. By virtue of the fact that the judiciary is bound to answer within the limits of the questions asked within the controversy it is tasked with adjudicating, courts do not reach the assessment of the proportionality (in the strict sense of the term) of a measure they already judged as not necessary. The impediment, however, is also of a logical nature; it would be hard to argue about the proportionality of an unnecessary measure: how can an unnecessary restriction upon a right be proportionate? Nevertheless, the nature of open data and of its relationship with personal data protection lends itself to a very basic, and yet crucial consideration about the proportionality of a ‘balance’ between open data and data protection: there cannot, by definition, be any. This proposition moves from the interaction between two orders of considerations: the first one about the nature of open data, the second one about the nature of data protection.

At its core, data protection is a system of checks and balances meant to regulate personal data protection by disciplining, to a certain extent, how personal data processing is to be carried out. Those checks and balances constitute a regulatory architecture meant to channel the activity of personal data processing, so that the power and information asymmetry that runs between data subjects and data controllers is mitigated by default. Open data tenets reject the kind of checks and balances in which data protection substantiates, requiring complete accessibility, usability, and interoperability for data to be considered open. It is a logical conundrum: the balance between the controller’s interests and the rights and freedoms of the data subject is achieved by constraining personal data processing activities through rules and principles. At the same time, information does not get to be qualified as open data if its accessibility, usability, or interoperability is somehow constrained. In a way, open data requirements reject balancing and proportionality – fostering unfettered accessibility and usability is its whole point – while the right to data protection, and in particular the secondary

---

1257 Cfr. CJEU, Österreichischer Rundfunk (n 853) §91: “If the national courts conclude that the national legislation at issue is incompatible with Article 8 of the Convention, that legislation is also incapable of satisfying the requirement of proportionality in Articles 6(1)(c) and 7(c) or (e) of Directive 95/46. Nor could it be covered by any of the exceptions referred to in Article 13 of that directive, which likewise requires compliance with the requirement of proportionality with respect to the public interest objective being pursued. In any event, that provision cannot be interpreted as conferring legitimacy on an interference with the right to respect for private life contrary to Article 8 of the Convention”.

1258 See Hustinx (n 663); Docksey (n 914). Similarly, Hijmans (n 782).

1259 Cfr. CJEU, Digital Rights Ireland (n 767) §54: “the EU legislation in question must lay down clear and precise rules governing the scope and application of the measure in question and imposing minimum safeguards so that the persons whose data have been retained have sufficient guarantees to effectively protect their personal data against the risk of abuse and against any unlawful access and use of that data”.

205
framework in which it substantiates, is meant to strike a fair balance between countervailing interests.\textsuperscript{1260}

To put it otherwise, when the rights and freedoms underlying open data compete with the right to data protection, a balancing exercise would never result in personal data being released as open data, even when the result of the balancing test favours openness over protection.\textsuperscript{1261} The very notion of balance, and the very idea of proportionality, require the pursuit of a middle ground. Such a middle ground would substantiate, concretely, in some form of access control or use restriction that – by definition – put the information within a shade in the openness spectrum other than open data.

Open data was originally meant to encourage the waiver of disposable rights, like intellectual property, and the removal of other barriers to data sharing and re-use, such as business confidentiality. There is no need to recur to balancing when the rightsholder voluntarily relinquishes control over information. Open data was never meant to permit wiggling out of regulatory frameworks whose obligations cannot be derogated from the rightsholder, and while the consent of the data subject is a valid legal basis for the removal of many obstacles to data sharing and re-use that derive from data protection law (e.g. the need to comply with the lawfulness principle) many of the rules and principles that embody the right to personal data protection cannot be waived by the rightsholder – while it is an individual right, data protection also aims at safeguarding collective interests.

Attempts at running the idea of balancing open data and data protection through a \textit{stricto sensu} proportionality test by following the kind of formal legal argumentation procedures mentioned above\textsuperscript{1262} is hindered by the fact that the assessment should be performed on an actual measure, rather than on a general idea. It can however still help highlight how a ‘balance’ between open data and data protection cannot be achieved through proportionality reasoning. The process of testing the proportionality of a measure where a principle, in our

\textsuperscript{1260} Cfr. CJEU, Asociación Nacional de Establecimientos Financieros de Crédito (ASNEF) (C-468/10) and Federación de Comercio Electrónico y Marketing Directo (FECEMD) (C-469/10) v Administración del Estado, Joined cases C-468/10 and C-469/10, 24 November 2011, ECLI:EU:C:2011:777, where the Court ruled that national rules that require data to be publicly available for the controller to be able to rely on its legitimate interest as a ground for processing, “thereby excluding, in a categorical and generalised way, any processing of data not appearing in such sources” (§ 49), were precluded. The Court considered (§46, 47) that, while nothing precluded EU Member States from establishing guidelines for the balancing exercise required by the legitimate interest of the data controller, national rules excluding (or allowing) the possibility to process certain categories of personal data by “hard-coding” the result of the balancing exercise were illegitimate. See also CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845) §62: “Article 7(f) […] precludes Member States from excluding, categorically and in general, the possibility of processing certain categories of personal data without allowing the opposing rights and interests at issue to be balanced against each other in a particular case. […] Member States cannot definitively prescribe, for certain categories of personal data, the result of the balancing of the opposing rights and interests, without allowing a different result by virtue of the particular circumstances of an individual case”.

\textsuperscript{1261} Cfr. CJEU, Volker und Markus Schecke and Eifert (n 921), where a measure much less restrictive for the right to personal data protection than open data has been deemed as not being proportionate by the Court. The Court stated (§81) that “(t)here is nothing to show that […] the Council and the Commission took into consideration methods of publishing information […] which would be consistent with the objective of such publication while at the same time causing less interference with those beneficiaries’ right to respect for their private life in general and to protection of their personal data in particular, such as limiting the publication of data by name relating to those beneficiaries according to the periods for which they received aid, or the frequency or nature and amount of aid received”.

\textsuperscript{1262} See s. 5.4.3. above; Alexy (n 931); Klatt and Meister (n 1046).
case personal data protection, is interfered with by another principle, which in this case would ostensibly be open data, consists in the comparison between their abstract weight, in light of the concrete intensity of their interference, modulated according to the empirical and normative reliability\textsuperscript{1263} given to the consequences of the compression of one principle by another competing principle.

As a first step, one should identify which principles compete. Data protection is inherently difficult to define,\textsuperscript{1264} and also instrumental to several other rights and freedoms, but it does have somewhat definite boundaries. Open data, on the other hand, is not a right or an interest per se, but rather a set of requirements expressed by data sharing policies that may uphold a wide array of different rights and freedoms, from good administration to freedom of business. A first idiosyncrasy is thus that while the right to data protection can integrate a principle on its own, as it is an autonomous (fundamental) right, the competing principle cannot be constituted by open data \textit{per se}. Granted, there are rights and freedoms underlying open data, and it is those rights and freedoms that constitute the competing principle.

But let us go past the issues pertaining to open data not being a right or freedom \textit{per se}, and pretend it could constitute a principle on its own. What would its abstract weight be? One could argue for the weight of open data to be extremely high: the broad range of rights and freedoms that can be served through open data sharing and re-use – from freedom of information to the right of access – are certainly fundamental in a democratic society. Yet, it would seem more appropriate to weight open data in light of its alternatives, rather than through a binary approach: the choice is not between open and closed, but between unfettered accessibility and usability and some form of access control or use restriction. From this perspective, if open data could be a self-standing principle, its abstract weight would derive from the variance stemming from elimination of the checks and balances of which data protection is made, rather than from the sheer differential between open and closed.

As for the abstract weight of data protection, it is the weight that a qualified fundamental right has. In this respect, it is interesting to note how \textit{actual} proportionality balancing in case law generally moves from a position where the abstract weight of the competing principles is the same – e.g. data protection and access to documents, or privacy vis-à-vis freedom of expression. That is not the case in the context of assessing the proportionality of a compression of data protection, a fundamental right, by virtue of open data, a set of requirements expressing a data sharing and re-use policy.

The intensity of the interference of a principle upon the other, as opposed to their abstract weight, is a test that must move from the concrete characteristics of the case at hand. It is thus fairly difficult to speculate about interferences without the context deriving from a concrete case to adjudicate or a legislative proposal to evaluate. Yet, there are still some considerations that can be drawn from considering the interference of a principle upon another from an abstract perspective. On the one hand, the interference that would derive from personal data being shared and re-used as open data is, from an abstract perspective, of the starkest nature. As explained above, a high number of the checks and balances that constitute personal data

\textsuperscript{1263} I.e. what we know that we know, what we know that we do not know, and the fact that we do not know what we do not know.

\textsuperscript{1264} See the previous chapter.
protection – from purpose limitation to security – are inherently incompatible with the requirements that information must have to be considered open data.\footnote{See s. 5.5.1 above.}

It would not be enough, say, to lower the degree of specificity required by the purpose specification and limitation principles, or to waive post-publication security requirements: being able to define personal data as open data without muddling the latter’s meaning as an autonomous concept would require a much broader and harsher set of derogations than what currently allowed by the GDPR. Granted, however, a concrete measure would have a much narrower scope, likely identifying specific categories of personal data and of data subjects and controllers,\footnote{Cfr. CJEU, Volker und Markus Schecke and Eifert (n 921, §110), where the Court ruled against the validity of an EU law norm mandating the publication of personal information “in so far as […] those provisions impose an obligation to publish personal data relating to each beneficiary without drawing a distinction based on relevant criteria such as the periods during which those persons have received such aid, the frequency of such aid or the nature and amount thereof”.} so that the concrete intensity of the interference would be somewhat circumscribed. The other side of the medal, yet again, lends itself to two possible interpretations: on one hand, one might see the interference of data protection as absolute, as its rules and principles do impede the sharing and re-use of information as open data. On the other hand, and again depending on the concrete circumstances of the case, the interference may be seen as relative: it is not open data being compressed, but the rights and freedoms that underlie it, and those rights and freedoms may be upheld through a less permissive data sharing policy than open data.

The (normative and empirical) reliability of the premises on which the balancing exercise rests modulates the weight to the principles at stake and the intensity of the interference of a principle upon the other. It derives from the fact that “(t)he more heavily an interference with a right weighs, the greater must be the reliability of its underlying premises”.\footnote{Alexy (n 931) 408. See also Klatt and Meister (n 1046) 11.} The reliability of the premises on which the \textit{stricto sensu} proportionality test is informed has to be ascertained concretely: there is no point in hypothesising without an actual case to be adjudicated nor a legislative proposal to be assessed. Yet, what has been discussed so far would arguably suggest that the premises on which the idea of having to seek a balance between open data and data protection by virtue of the emergence of the ‘smart city’, both from a normative and an empirical perspective, are shaky at best. The idea that open data is necessary for smart cities to develop, and that some undefined categories of personal data would be amongst such necessary data, and that no data sharing policy other than open data would do, and that personal data protection is in the way of its processing seems to be supported by an aspirational agenda, rather than from empirical evidence or specific cases or scenarios.

All in all, open data would seem to be, rather than a standalone principle, the outcome of a balancing exercise where either the weight of a principle (in this case data protection) or the intensity of the interference upon it is null, which seems far from being the case. In other words, open data is not a right, nor a principle, but a set of requirements that express a data release policy which may be used to uphold other rights or principles. When those rights and principles compete with the right to personal data protection, a \textit{stricto sensu} proportionate outcome cannot – by default – be open data sharing and re-use.\footnote{“No automatic priority can be conferred on the objective of transparency over the right to protection of personal data […] even if important economic interests are at stake” – CJEU, Volker und Markus Schecke and} A balance between
openness and protection entails the pursuit of a middle ground that would, as such, fall short of the unfettered accessibility and usability requirements necessary to call information ‘open data’.

The open data movement was born out of concerns about other rights and freedoms, such as intellectual property and business confidentiality, where the agreement of the rightsholder can elide the interference upon the rights that could be compressed by publishing and re-using information as open data. Data protection is different: some of the rights and obligations stemming from it can be waived by the rightsholder, but many (if not most) of its rules and principles cannot be waived by data subjects. As a consequence, differently from what can happen e.g. with intellectual property rights, the interference upon data protection cannot be excluded entirely, and thus the result of the balancing exercise between data protection and the rights and freedoms that can be advanced through information sharing and (re)use cannot possibly be open data.1269

5.5.5 DOES IT RESPECT THE ESSENCE OF THE RIGHT?

The preceding chapter, in framing this thesis’ theory of data protection’s substance and essence, underscored how both the concept essence of fundamental rights in general,1270 and the essence of the right to personal data protection in particular,1271 are still highly debated in European doctrine and jurisprudence. Several authors, and some CJEU judgements, provided different interpretations of what constitutes the essence of data protection as a standalone fundamental right, and the previous chapter of this thesis did too. This section argues that the idea of releasing personal data as open data is at odds with all of them.

This thesis’ theory of data protection frames it as a system of checks and balances1272 that embodies a collective posture towards the promises and perils of personal data processing—a societal choice towards the regulation of computing that gradually materialised into a multifaceted system of rules and principles. The thesis constructs data protection as a permissive right, where an interference upon it is not given by the mere processing of personal data, but rather from processing activities that deviate from the rules and principles set by secondary legislation. Accordingly, it conceives of the essence of data protection by moving from its interferences, and defines as an interference upon the essence of the right a compression that, by challenging the functioning of such system of checks and balances in its entirety, defies the collective choice it underlies. Open data, in its categorical rejection of any sort of restriction upon information access and use, is expression of a diametrically opposed kind of decision – one aiming at freeing information, rather than channelling it. From this

Eifert (n 921) §85. See also CJEU, ClientEarth and Pesticide Action Network Europe (PAN Europe) v European Food Safety Authority, C-615/13 P 16 July 2015, ECLI:EU:C:2015:489, §51. 1269 Cfr. CJEU, Manni (n 1211) §57 ss, where the Court, when dealing with the proportionality of an interference upon data protection, notes that the measure under scrutiny in the case at hand, which mandated the publication of individuals’ personal data, still required the disclosure “only for a limited number of personal data items”, and that “it appears justified that natural persons who choose to participate [...] are required to disclose the data relating to their identity and functions [...] especially since they are aware of that requirement when they decide to engage in such activity”.

1270 Dawson, Lynskey and Muir (n 934); Lenaerts (n 924); Tridimas and Gentile (n 926).

1271 Porcedda (n 918); Brkan, ‘The Concept of Essence of Fundamental Rights in the EU Legal Order: Peeling the Onion to Its Core’ (n 930); Brkan, ‘The Essence of the Fundamental Rights to Privacy and Data Protection: Finding the Way Through the Maze of the CJEU’s Constitutional Reasoning’ (n 939).

1272 See similarly Hustinx (n 663); Docksey (n 914); Hijmans (n 782).
perspective, the idea of ‘balancing’ open data and data protection clashes with the latter’s very essence.

The pursuit of a balance between open data and data protection also seems at odds with several other framings of the essence of the right to personal data protection that emerged in doctrine and jurisprudence. Lynskey postulates that the essence of the right might be sought in data security – the ‘technical and organisational measures’ that permeate secondary data protection legislation and that, despite being absent amongst the elements listed in Art. 8 of the Charter, flows from its interpretation.\(^\text{1273}\) The requirements underlying open data reject any kind of security measure successive to the information’s publication. The information’s integrity, confidentiality, and availability can be safeguarded before its publication as open data; afterwards, it can be accessed and used by anyone for any purpose, which is the exact opposite of confidential, and duplicated at will and in a machine-readable and interoperable format, which is ideal for its availability, but at odds with the original dataset’s integrity. If the essence of data protection is to be sought in security, the idea of releasing personal data as open data is in outright conflict with it.

A number of other authors frame the inviolable core\(^\text{1274}\) of data protection as comprised by several of its most essential elements, generally the FIPPs with the addition of some extra element that gained prominence over time. Tzanou, for instance, submits that “(t)he core essence of the right to data protection is laid down in Article 8 […] Furthermore, sensitive data […] should be shielded from certain categories of processing […] The purpose specification principle should also have a ‘hard core’”.\(^\text{1275}\) Porcedda, similarly, extracts the essence of data protection\(^\text{1276}\) by moving from a list of its core attributes as emerged in law, jurisprudence, and doctrine, and that is (partly) overlapping with the elements listed in Art. 8 of the Charter. Bygrave, while not dealing directly with the concept of essence\(^\text{1277}\) of data protection in the EU legal order, discusses the “core principles” of international data privacy law, again falling back to the FIPS plus some additional elements.\(^\text{1278}\) This ‘positivist’ doctrinal view of the essence of data protection, regardless of whether it relies only on the elements listed in Art. 8 of the Charter or on the evolution of the FIPPs, always include elements of data protection that, by somewhat curtailing the possibility to process personal data without any constraint, are fundamentally at odds with open data requirements. The purpose limitation principle, for instance, is both a cardinal element of data protection – part of its essence – and diametrically opposed to the open data ethos: ‘serendipitous re-use’ and the specification and limitation of the processing’s purposes are mutually incompatible. If the essence of the right to personal data protection is to be sought in the core principles in which it substantiates – e.g. the FIPPs, or the ones contained in Art. 8 of the Charter, or even just the ones explicitly mentioned by the CJEU – then the idea of publishing and re-using personal data as open data would encroach upon it.

\(^\text{1273}\) Lynskey, The Foundations of EU Data Protection Law (n 19) 271.
\(^\text{1274}\) Or “essence”, or “essential traits”, depending on the wording used. See generally Dawson, Lynskey and Muir (n 934); Lenaerts (n 924); Tridimas and Gentile (n 926).
\(^\text{1275}\) Tzanou (n 649) 97, 98. See also Maria Tzanou, The Fundamental Right to Data Protection: Normative Value in the Context of Counter-Terrorism Surveillance (Bloomsbury Publishing 2017).
\(^\text{1276}\) Porcedda (n 918).
\(^\text{1277}\) The author indeed specifies that “(i)t bears emphasis that the rules giving effect to these principles are seldom set down as absolutes”, and it is thus questionable whether they constitute actual ‘essential traits’: Bygrave, Data Privacy Law: An International Perspective (n 269) 145.
\(^\text{1278}\) Such as “sensitivity” – Bygrave, Data Privacy Law: An International Perspective (n 269) 165.
Some still hold that the essence of data protection is privacy; others seek it in individual control or informational self-determination. Privacy is an extremely multifaceted notion by nature, and data protection as informational self-determination is a questionable construction in light of the evolution of data protection legislation after (and on the model of) Convention 108. Regardless, the idea of sharing and re-using information relating to identifiable individuals as open data is intuitively irreconcilable with any sort of reasonable expectation of privacy or ability to control the flow of information by the data subject. The extent of the mental gymnastics needed to reconcile open data, on one hand, and privacy and self-determination, on the other, seems beyond reasonable. If privacy or individual control were at the essence of the right to data protection, ‘balancing’ them with open data would require compromising on their meaning.

The essence of the EU right to personal data protection is still, to a large extent, unclear. There are several distinct interpretations of what data protection’s substance and essence may be. This thesis moves from the construction of data protection as a system of checks and balances, and of its essence as the collective decision of regulating personal data processing, which is in essential conflict with the rationale underlying open data. Other authors take different stances, and seek the essence of the right in data security, in the FIPPs, or in data protection’s roots in privacy and informational self-determination. The idea of ‘balancing’ open data and data protection, in any event, is in unambiguous contrast with all of them.

5.6 CONCLUSION

The idea of balancing data protection with open data is not really about balance, but rather about de-regulation. Personal data processing must be subject to at least some rules and principles, and the constraint deriving from compliance with those norms are in stark contrast with the requirements that must be satisfied for information to be called ‘open data’. Personal data, in other words, cannot be open data: unless data protection law is neglected, tertium non datur. That does not mean that personal data cannot be shared or used tout court, or that secondary purposes are not allowed at all, or that the data subject’s involvement is necessary for the processing to be legitimate – all common misconceptions – but just that a strict definition of open data is based on requirements that are in fundamental contrast with data protection’s law’s main tenets.

Such conflict is recognized, at a legislative level, in all the instruments that, directly or incidentally, promote open data publishing and re-use in the EU, which always contain exceptions and non-prejudice clauses on personal data protection grounds. But even if it were not possible to subsume the relationship between open data and data protection under a clear (set of) rule(s) – if we had to balance open data and data protection through the lenses of lato sensu proportionality by assessing the suitability, necessity, and proportionality stricto sensu of a measure – the result of such balancing test could never be entirely open data. Even if it were provided by law and suitable to the achievement of objectives of general interest, which

---

1279 E.g. Poulet (n 881). To some extent, Hijmans (n 782).
1280 See generally Lynskey, The Foundations of EU Data Protection Law (n 19) 177 ss.
1281 A line of thought originating with the German “census decision” – see Hornung and Schnabel (n 1204).
1282 Defining private life has been held as both impossible and unnecessary by the ECtHR in several occasions: see e.g. Niemietz v. Germany (n 300) §29; Peck v. United Kingdom (n 300) §57.
1283 See Hustinx (n 663).
is likely – and yet not granted – pursuing a ‘balance’ between open data and data protection is not necessary, given that data protection is already a permissive regime as it is. It would also not be proportionate, since open data, being at the most extreme end of the openness spectrum, rejects proportionality by definition. It may indeed be argued that, if data protection does have an essence, the idea of balancing it with open data would require encroaching upon it.

Yet again, that does not really preclude the pursuit or the achievement of whichever ‘smart city’ objective a municipality might set. Data protection law does not actually impede data sharing and re-use, nor the realisation of the rights and freedoms underlying open data: it is just opposed to the unfettered lack of restrictions open data requirements are based on. There are two ways to see the issue. One is just to see data protection as an insurmountable barrier for open data; the other way is to see the protection as personal data, and compliance with data protection legislation, as an additional, negative, and implicit requirement of open data. Regardless, the result is the same: if it is personal data, and is processed in compliance with the law, it cannot be open data. Arguing otherwise – pursuing a ‘balance’ between open data and data protection even when the former’s requirements embody an outright rejection of balancing and proportionality – is tantamount to arguing that personal data processing should be deregulated, which goes beyond the realm of the justiciable and into policy, and is also a policy debate we ought not to have.

That is not to say that there are no issues in the current regime, but rather that the solution of those issues is not to be sought in balancing open data and data protection – at least, not in ‘balancing’ as generally understood by legal scholarship. Secondary data protection legislation is the framework with which data subjects’ rights and freedoms, on one hand, and the rights and freedoms on which open data is based, on the other, are balanced. The result of such balancing, when personal data are in play, cannot be open data, as open data is not proportionate by definition. As mentioned, however, there is an aspect about the interaction between the instrumentation and datafication of the built environment (i.e. the emergence of the ‘smart city’), data protection, and open data that legitimises the concerns from which this thesis moves from. Specifically, the notion of personal data and the material scope of EU data protection law.

The technological landscape and the state of the art in data processing has been rapidly changing over the years that passed since EU data protection law first came to be. The amount and granularity of the data available has been steadily increasing, and the costs of storing and processing it have been progressively decreasing. Expertise and know-how are getting increasingly widespread, too, which is connected with a growing ease in singling out individuals from (purportedly) anonymised records, and in drawing inferences about people’s attributes and characteristics from (ostensibly) unrelated data. At the same time, the interpretation of what constitutes personal data – any information relating to an identified or identifiable natural person – has been constructed very extensively by doctrine.

---

1284 See e.g. Mayer-Schönberger and Cukier (n 13).
1285 See Ohm (n 42).
1286 See Ohm and Peppet (n 1254).
1287 GDPR, Art. 4(1).
1288 E.g. Purtova, ‘The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law’ (n 903); van Loenen, Kulk and Ploeger (n 977); Kulk and Van Loenen (n 39).
jurisprudence, and supervisory authorities alike. As a consequence, it may be reasonably feared that the expansion of the notion of personal data in EU data protection law could grow unchecked, and begin to encompass within its scope also information traditionally seen as non-personal, and thus safe for release as open data. Yet again, that has nothing to do with balancing, but rather pertains to the development of a nuanced and reasonable interpretation of the concept of personal data, and will be discussed in the following chapter.

---

1289 CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845); Peter Nowak v Data Protection Commissioner (n 845).

6. THE TROUBLE WITH PERSONAL DATA

The previous chapter outlined how the idea of balancing open data and data protection is not really about balancing nor proportionality, but rather a policy stance towards the (de)regulation of personal data processing. Open data’s constitutive traits, by definition, reject the kind of balancing that is bound to take place when individuals’ right to personal data protection comes into play. The core of the matter is not the pursuit of the impossible balance between open data and personal data protection, but rather the definition of the boundaries of the material scope of data protection legislation: personal data. The concept of personal data, defined as ‘any information relating to an identified or identifiable natural person’, is core to the European data protection framework since its very inception.\(^\text{1291}\) The qualification of data as personal is a *conditio sine qua non* for processing to be considered within the material scope of the General Data Protection Regulation (GDPR), and thus for the applicability of EU data protection law.\(^\text{1292}\) Despite the crucial importance of the notion, its boundaries are however oftentimes blurry. Ascertaining whether data is personal frequently depends on each individual processing’s context and characteristics. The qualification of data as personal is contextual: the same piece of information can be anonymous in one moment in time and personal in another.\(^\text{1293}\) As a result of the relative character of the notion of personal data, much is left to the discretion of the interpreter.

As open data and personal data are meant not to overlap, the focal point of their interaction becomes determining what is personal and what is anonymous, for at least two reasons. The first one is legal clarity: both personal data controllers and open data stakeholders would benefit from a clearer definition of what is personal and what is anonymous. The second reason is scope creep: the more kinds of information are considered to be personal data, the fewer can be released as open data. Indeed, doctrine and jurisprudence appear to favour an expansive interpretation of the notion of personal data, setting a very low identifiability threshold.\(^\text{1294}\)

Thus, on one hand, according to the GDPR,\(^\text{1295}\) identifiability must be ascertained taking into account every means reasonably likely to be used by the data controller or by another person. On the other hand, the suitability of anonymization as a protection mechanism, and of anonymity as a basis for policy, have been subject to intense doctrinal criticism,\(^\text{1296}\) especially

---


\(^\text{1292}\) GDPR, Art. 2(1).


\(^\text{1294}\) CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845).

\(^\text{1295}\) GDPR, Recital 26.

following a number of successful re-identification attacks performed on (purportedly) anonymised datasets.

At the same time, the ways in which the information can be said to be relating to a natural person are manifold: information can relate to a natural person not only by virtue of its content, but also due to the purpose of the processing, or of its result. Depending on each individual processing instance, all data can thus potentially become personal. The combination of the low identifiability threshold and the wide range of ways in which data can relate to a person can lead to an extremely wide material scope of EU data protection legislation. Data protection can become, it has been argued, ‘the law of everything’, used to tackle problems which it was originally not meant to solve.

This chapter responds to the growing concerns surrounding the perceived over-inclusiveness of the notion of personal data, highlighting the elements that can be used to develop a balanced approach to its interpretation. The concept of personal data was drafted to be broad and technology-neutral enough to avoid leaving any personal data processing instance unprotected. Its boundaries are elastic, and their extension depends on the reading given by the interpreter, and eventually by the courts. Ultimately, doctrine and jurisprudence bind the reading of the law, and concur in setting the concrete extension of the concept of personal data. I argue that, while the notion of personal data lends itself to the possibility of an overly expansive interpretation, there is also room to construe it narrowly enough to withstand the tests to come. While investigations on the tenability of the notion of personal data as a regulatory instrument are certainly necessary and welcome, the GDPR has just started being applicable and is expected to remain in force for quite some time. It seems thus necessary, in consideration of the concern about the breadth of the notion of personal data, to also highlight the elements that can prevent its overextension in the short run. Several authors pointed out the inconsistencies of the concept of personal data, sometimes advocating for a regulatory

---

1297 This chapter defines attacker as the subject attempting to identify or re-identify an individual within a record, and the attack the activities performed by the attacker for such purposes.
This chapter starts by defining the concept of personal data under EU data protection law. It then delves into the critiques raised towards its two most crucial elements: identifiability, and the link between information and natural person. The chapter follows up by providing the elements for a balanced reading of the notion of personal data, pleading for a nuanced approach in its interpretation so that it does not unduly limit the possibility to share and re-use information as open data, and conversely by considering how open data can interfere with individuals’ privacy and data protection even if (purportedly) anonymised. To achieve further clarity in the definition of what falls under the material scope of EU data protection law and what does not, it continues by looking at how data protection might learn how not to deal in absolutes from the fields of information security and statistical disclosure control, i.e. by starting to consider anonymity as a process, rather than a state. This chapter concludes with a reflection about the meaning of the interpretation of the notion of personal data it offers for the relationship between open data and personal data protection.

6.1 PERSONAL DATA IN EU DATA PROTECTION LAW

Personal data is defined, within the EU data protection framework, as ‘any information relating to an identified or identifiable natural person’. An identifiable natural person is one that can be identified, directly or indirectly, by reference to a direct identifier, such as a name or an identification number, or by a combination of indirect identifiers, e.g. location data, IP addresses, or other factors specific to her identity. As the Article 29 Working Party (A29WP) highlights, there are thus four core components to the notion of personal data: information, a natural person, identifiability, and a link (‘relating to’) connecting the information and the data subject. Those components are cumulative: each is necessary to qualify data as personal. Lacking one or more of those components (e.g. the identifiability of the natural person to which the information relates, or the ‘relating to’ link between information and natural person) the data is not personal. The boundaries of the concept of personal data depend on the interpretation of, and the relation between, those four concepts.

The notions of information and of natural person as defined in EU data protection law will be examined more briefly than the concept of identifiability, or the link connecting data and data subject, since their legal construction within the current legal framework makes them less prone to be modulated by interpretative means. Any kind of information, regardless of its nature, content, format or the medium in which it is contained, can qualify as personal. It does
not need to be truthful or objective, nor secret or private, nor kept in a particular format or medium. Any sort of data can be personal, if it relates to an identifiable natural person: EU data protection law constructs the concept of information (or data, which it treats as synonyms) as broadly as possible. What constitutes a natural person, on the other hand, does not seem to raise particular problematics in respect to the extension of the material scope of data protection law, and is mostly left for MS law to determine.

The plasticity of the notion of personal data within the current legal framework derives mostly from the possibility to interpret identifiability and the meaning of ‘relating to’, rather than the (all-encompassing) view of information adopted by EU data protection law, or the qualification of what constitutes a natural person.

6.1.1 ‘RELATING TO’

The data and the natural person must be connected by a link: the information must be relating to the data subject. That does not necessarily mean that the content of the information must identify the data subject. The wording ‘relating to’ delineates a broader range of ways in which information can connect to a person. There are indeed three ways in which data can relate to the person: through its content, but also through its purpose, or the result of its processing. Those ways are alternative: one suffices to link the data and the natural person.

Information relates to a data subject when its content is about that specific natural person, regardless of its actual purpose or impact, ‘the most obvious and common understanding in a society of the word ‘relate’”. An identity card, for instance, is personal data in that its content is about a natural person, to which it links irrespective of the purpose or result of the processing; likewise, a medical analysis relates to the patient by virtue of its content, notwithstanding the purpose or result of the processing.

Data can also relate to the data subject when, despite not being about him or her by virtue of the content element, it is used or likely to be used with the purpose of evaluating, influencing, or generating consequences for that natural person. The readings of the accelerometer in a smart phone, despite being about the phone, can very well become personal data, for instance.

---

1308 ‘(A)s the law stands, and depending on the theoretical perspective towards the meaning of information, everything can still be plausibly argued either to be or to contain information that can be personal data provided the other requirements of the definition are met’: Purtova, ‘The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law’ (n 903) 27.
1309 The GDPR does not apply to the personal data of deceased persons; however, Member States are allowed to regulate with national law the processing of personal data of deceased persons (GDPR, Recital 27). On the topic, see Edina Harbinja, ‘Does the EU Data Protection Regime Protect Post-Mortem Privacy and What Could Be the Potential Alternatives’ (2013) 10 SCRIPTed 19.
1310 Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40) ss. 9 ss. See also CJEU, Peter Nowak v Data Protection Commissioner (n 845) which expands the view taken in CJEU, YS and others (n 845).
1312 Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40) 10. See also CJEU, Peter Nowak v Data Protection Commissioner (n 845) §37; 43.
when processed by a fitness app, ‘with the purpose to evaluate, treat in a certain way or influence the status or behaviour of an individual’.\textsuperscript{1313}

The result element links the information and the data subject where, despite not being about that natural person, nor meant to be used with a particular purpose concerning her, the data will have, or is likely to have, an impact or effect on the data subject. It is not necessary for the impact to be major: ‘It is sufficient if the individual may be treated differently from other persons as a result of the processing of such data’.\textsuperscript{1314} The result of the processing, whichever the content or the original purpose of the information, can thus render any kind of data personal. Asset monitoring through sensors involves data whose content is about the asset monitored, processed with the purpose of monitoring that asset. However, if such assets are entrusted to a natural person by her employer, then the data could arguably be used with the additional purpose of evaluating the employee’s performance, and would thus be considered as relating to the employee.

The GDPR’s text does not explicitly construe the ‘relating to’ link as being integrated by content, purpose, and/or result. The currently\textsuperscript{1315} dominant interpretation of what ‘relating to’ means derives from the A29WP Opinion on the concept of personal data.\textsuperscript{1316} While the A29WP’s Opinions are not binding, they do carry a large degree of significance in EU data protection doctrine and practice. The Court of Justice of the European Union (CJEU), the chief judicial authority of the EU, tasked with ensuring uniform interpretation of EU law, itself implicitly adhered to the construction of the relational tie set forth by the A29WP’s Opinion on the concept of personal data in its YS and Nowak judgements,\textsuperscript{1317} and explicitly referred to another A29WP opinion in the recent Jehovan todistajat case.\textsuperscript{1318} National courts and supervisory authorities consider them in their proceedings, too. The A29WP’s interpretation of the wording ‘relating to’, and of the notion of personal data tout court, is thus prominent in the European data protection milieu, cemented by the CJEU’s interpretation in the YS and Nowak cases.

It is hence not only the content that qualifies the information as relating to a natural person, but also the purpose of its processing, or its (likely) result. To integrate the definition of

\textsuperscript{1313} Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40) 10. See also Peter Nowak v Data Protection Commissioner (n 845) §38; 43.
\textsuperscript{1314} Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40) 11. See also Peter Nowak v Data Protection Commissioner (n 845) §39; 43.
\textsuperscript{1315} While the A29WP’s Opinion refers to the concept of personal data in the Data Protection Directive, the notion did not change with the enactment of the GDPR (see Opinion of AG Kokott in case Peter Nowak v Data Protection Commissioner, C-434/16, 20 July 2017, ECLI:EU:C:2017:582, §3), and what the Opinion holds is expected to remain valid: Purtova, ‘The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law’ (n 903) 6.
\textsuperscript{1316} Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40) 9 ss.
\textsuperscript{1317} CJEU, YS and others (n 845); Peter Nowak v Data Protection Commissioner (n 845).
\textsuperscript{1318} CJEU, Jehovan todistajat (1122) §21: ‘it is clear from Opinion 1/2010 of 16 February 2010 on the concepts of ‘controller’ and ‘processor’ produced by the Working Group set up pursuant to Article 29 of Directive 95/46, that, in particular, the ‘effective control’ and the conception that the data subject has of the controller must be taken into account’. While case C-25/17 is to my knowledge the first time where the Court referred to an A29WP opinion explicitly, implicit references can be found in other cases, e.g. YS and others (n 845), Nowak (n 845), and Google Spain (n 1118). Furthermore, several Opinions of the Attorney General, which have been relied upon by the Court, explicitly refer to A29WP opinions.
personal data, and thus to trigger the applicability of the GDPR, the natural person to which the information relates must however be also identified or identifiable.

6.1.2 IDENTIFIED OR IDENTIFIABLE

A natural person is considered identified when isolated from a group of reference, and identifiable when it has not, despite the concrete possibility to do so. EU data protection law applies in both cases: the mere possibility to identify the natural person to whom the information refers is sufficient to trigger the applicability of the GDPR, it is not necessary for the data subject to be identified already.\footnote{\textsuperscript{1319} GDPR, Art. 4.1; Recital 26. See also Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40) 12.}

Data subjects can be identifiable both directly or indirectly. Identification is performed through pieces of information commonly called ‘identifiers’,\footnote{\textsuperscript{1320} For an original taxonomy on identifiers in data protection, see Ronald Leenes, ‘Do They Know Me? Deconstructing Identifiability’ (2008) 4 University of Ottawa Law & Technology Journal 135. The definition of identifiers this chapter relies on is taken from ISO/TS 25237:2008 (Health informatics — Pseudonymization).} which can be either direct or indirect. Direct identifiers are data that identifies a single individual, either without additional information (e.g. my employee ID number) or by cross-correlating it with other information.\footnote{\textsuperscript{1321} See CJEU, Lindqvist (n 845) §27.} Indirect identifiers are data that does not allow the identification of the data subject on its own, but can reduce the sample to which the data subject belongs until, by correlating enough indirect identifiers, he or she becomes unique within the record and can thus be singled out.\footnote{\textsuperscript{1322} ‘The use by the EU legislature of the word ‘indirectly’ suggests that, in order to treat information as personal data, it is not necessary that that information alone allows the data subject to be identified’: CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845) §41. See also Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40) 13.}

According to the GDPR, to establish identifiability, account should be taken of all the means reasonably likely to be used to identify the data subject, directly or indirectly. To ascertain whether such means are reasonably likely to be used, the interpreter should consider all objective factors inherent to the processing, considering both the available technology at the time of the processing and the foreseeable technological developments.\footnote{\textsuperscript{1323} GDPR, Recital 26; Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40) 15.} Such means, furthermore, do not need to be used directly by the controller to be relevant for the identifiability test, but can be employed by either the data controller or by another person as well.\footnote{\textsuperscript{1324} GDPR, Recital 26; CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845) §43; Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40).} The wording of Recital 26 ‘suggests that, for information to be treated as ‘personal data’ […] it is not required that all the information enabling the identification of the data subject must be in the hands of one person’.\footnote{\textsuperscript{1325} CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845) §43.}

6.1.3 ANONYMITY IN DATA PROTECTION

Anonymous data in EU data protection law can be defined, 	extit{a contrario}, as non-personal data; it refers to information which does not relate to an identified or identifiable natural person (anonymous data \textit{stricto sensu}), or to information that has been processed to ensure that the
natural person to which it refers is not identified or identifiable anymore (i.e. anonymized data). Anonymous data are indeed not encompassed within the material scope of European personal data protection legislation: their processing does not trigger the applicability of data protection principles.\textsuperscript{1326} There are two ways in which information can be considered anonymous: when it does not relate to a data subject, or when the data subject it relates to is not identified or identifiable. The determination must be performed on a case-by-case basis, and depends on the concrete characteristics of each individual processing instance. Anonymity is inherently contextual and relative.\textsuperscript{1327}

Data is anonymous when the identifiability requirement is lacking, i.e. when the natural person it refers to is not identified or identifiable. The capacity to identify the individuals to which the information relates must be modelled in consideration of all the means reasonably likely to be used, either by the data controller or by any other person, to identify (directly or indirectly) the natural person to which the information refers.\textsuperscript{1328} Determining whether means are reasonably likely to be used to identify the data subject should be done in consideration of all the concrete, objective factors pertaining to the processing instance at hand,\textsuperscript{1329} taking into account both the state of the art at the time of the processing and future technological developments. Anonymous information can, over time, become personal data, due to technological advances and increase in the available auxiliary information.

Data can be anonymous \textit{stricto sensu}, from the beginning of its lifecycle, or it can become anonymous afterwards, when it is anonymized \textit{ex post}. Anonymization is the processing of personal data\textsuperscript{1331} which aims at preventing, irreversibly, the identification of the data subject. The solidity of each anonymization instance – and hence the personality or anonymity of the information concretely processed – can be evaluated considering the capacity of a hypothetical

\textsuperscript{1326} GDPR, Recital 26: “[…] The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes”.

\textsuperscript{1327} See Enrico Pelino, ‘La Nozione Di Anonimo’ in Giusella Finocchiaro (ed), \textit{Diritto all’anonimato: anonimato, nome e identità personale}, vol 48 (CEDAM 2008).

\textsuperscript{1328} GDPR, Recital 26. See also DPD, Recital 26; CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845).

\textsuperscript{1329} E.g. the costs of and the amount of time required for identification, the expertise and motivation of the attacker, etc.


\textsuperscript{1331} Indeed, the act of anonymising personal data constitutes an instance (further) processing of those personal data, with all the consequences that derive from it (e.g. the necessity to comply with the purpose limitation principle).
attacker to single out an individual, link records relating to an individual, or infer a value or attribute about an individual moving from other values or attributes.

6.1.4 Pseudonymity in data protection

Anonymity is meant to be irreversible within the state of the art available at the time of processing. In that it differs from pseudonymity, which is the processing of personal data, generally by substituting a direct identifier or a unique attribute with a pseudonym, so that the information can no longer be linked to the data subject without the use of auxiliary data. Contrary to anonymous or anonymized data, pseudonymous data is still data about an (indirectly) identified or identifiable natural person, and hence personal under EU data protection law. Pseudonymity must be ensured through technical and organisational measures aiming at keeping the pseudonymous dataset separate from the auxiliary information necessary to identify the data subjects the information refers to. It has been noted how the separation between the pseudonymous dataset and such auxiliary information is a key component of the definition of pseudonymity under EU data protection law.

Pseudonymous data does not represent a new category of information, a tertium genus between personal and anonymous data. It is not, unlike e.g. sensitive data or “Art. 11 data”, a category of information subject to a different regime, either. Pseudonymity is, rather, a prime example of those technical and organisational measures that are ubiquitously present in secondary data protection legislation. Pseudonymization is addressed, for instance, in Art. 6(4) of the GDPR, which indicates it amongst the factors to consider when assessing the compatibility (rectius, non-incompatibility) of an instance of further processing vis-à-vis the original collection’s purposes. It is also mentioned in both Art. 25(1) and 32(1) on data protection by design and on the security of the processing, respectively, as a way to implement data protection principles and to heighten the security of the processing, and by Art. 89(1) as one of the safeguards that may be used to balance the leniency of the regime applicable to

---

1332 Linkability can be further split between local linkability (the ability to link different records pertaining to the same data subject within a single dataset), domain linkability (the ability to link different records pertaining to the same data subject within a two or more datasets available to the data controller), and global linkability (the ability to link different records pertaining to the same data subject within a two or more datasets, regardless of whether available to the controller or not) – Runshan Hu and others, ‘Bridging Policy, Regulation, and Practice? A Techno-Legal Analysis of Three Types of Data in the GDPR’ in Ronald Leenes and others (eds), Computers, Privacy and Data Protection: the Age of Intelligent Machines (Hart Publishing 2017) 126.

1333 Article 29 Data Protection Working Party, ‘Opinion 05/2014 on Anonymisation Techniques WP216’ (n 521) 10. See also Hu and others (n 1332).


1335 GDPR, Art. 4(5): “pseudonymisation’ means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person”.

1336 GDPR, Recital 26: “[…] Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person”. See also Article 29 Data Protection Working Party, ‘Opinion 05/2014 on Anonymisation Techniques WP216’ (n 521) 20.

1337 Hu and others (n 1332) 127.

1338 See the following section.
personal data processing for archiving, statistical, and scientific or historical research purposes with individuals’ right to personal data protection.

6.1.5 “Art. 11 Data”

It may be argued that the GDPR introduces, practically if not explicitly, a new category of information, which can be referred to as “Art. 11 data”. Art. 11(1) of the GDPR states that, if the purposes of the processing do not require the identification of the data subject, the controller is not obliged to maintain, acquire, or process additional information to identify the data subject for the sole purpose of complying with the GDPR. If the controller can demonstrate that it is not in a position to identify the data subject, Art. 11(2) waives the applicability of a large number of data subject rights, except where the data subject, for the purpose of exercising his or her right, provides additional information enabling his or her identification.

The rationale underlying Art. 11 is that information can be personal even if the data subject cannot be identified by the controller alone, despite being identifiable by entities other than the controller. In those cases, (re)identifying the data subject for the sole purpose of complying with the obligations deriving from the exercise of his or her rights would be unduly burdensome for the data controller, and thus Art. 11 waives those obligations. If, however, the data subject is able to provide the controller with auxiliary information that makes the identification seamless, the costs, time, and effort necessary for the controller to identify the data subject ceases to overweight the data subject’s interest in the exercise of his or her rights, and the Art. 11 regime ceases to apply.

Art. 11 can thus be exploited by data controllers to lessen the burden of compliance, provided that their processing architecture is engineered accordingly. The applicability of the regime ex Art. 11 is indeed conditional to the lack of local and domain linkability – the ability to identify a data subject within a single dataset or through the combination of two or more datasets in possession of the controller. Art. 11 represents a fitting example the kind of mechanic that substantiate the GDPR’s ‘risk-based approach’, as it allows data controllers to modulate the intensity of their data protection obligations by structuring their processing activities as to reduce data subjects’ identifiability to the minimum necessary for the achievement of the processing’s purposes. By providing the data controllers with incentives (i.e. fewer obligations), Art. 11 fosters compliance with fundamental data protection principles and practices, such as data minimisation and storage limitation.

1339 See e.g. Hu and others (n 1332).
1340 A circumstance about which the data subject must be informed accordingly, if possible.
1341 I.e. Art. 15 to 20 of the GDPR.
1342 Cfr GDPR, Recital 26. A natural person that, as a private individual, operates a website that logs visitors’ IP addresses for security reasons is arguably not in the position to identify the data subjects to which those addresses relate – but ISPs can. IP addresses are thus still personal data even when in the hands of the operator of a website, but (as the operator, as a private individual, is not in the position of identify such information on its own) the regime ex Art. 11 applies.
1343 Hu and others (n 1332) 128. Art. 11 is however unconcerned with global linkability, i.e. the ability to link two or more datasets that can be in possession of different data controllers.
1344 Article 29 Data Protection Working Party, ‘Statement 14/EN WP 218 on the Role of a Risk-Based Approach in Data Protection Legal Frameworks’ (n 1226).
Art. 11, just like pseudonymous data, does not challenge the anonymity/identifiability dichotomy on which the material scope of EU data protection law rests. It does not introduce a middle ground between anonymous and personal, nor represents a separate regulatory regime for information that refers to an identifiable (as opposed to identified) natural person, either.\footnote{As wished e.g. in Schwartz and Solove (n 1302); Solove and Schwartz (n 1302).} Art. 11 data is still just personal data.

Making personal data into Art. 11 data does not make them anonymous – even if the data controller is not in a position to identify the data subject anymore, third parties might still be able to do so. Art. 11 just provides for a more lenient regime towards personal data relating to persons that are identifiable by an entity other than the controller, which might help address the tension inherent to the relationship between open data and data protection by weighting down the burden of compliance with data subject rights.

### 6.2 The Potential Over-Inflation of the Concept of Personal Data

The concept of personal data, as summarised above, is very wide;\footnote{CJEU, Rijkeboer, C-553/07, 7 May 2009, EU:C:2009:293, §59.} its definition is broad and contextual, highly dependent on the reader’s interpretation. It is also largely technology-neutral, in that it is applicable to any kind of data type and processing technique, and covers situations where the identification of the data subject is merely potential (e.g. when the natural person to whom the information refers is indirectly identifiable through the means available by a person other than the controller).

Several authors have thus pointed out, directly or indirectly, the deficiencies of the notion of personal data, highlighting how its current normative, doctrinal, and judicial construction may backfire in the near future, overly inflating the material scope of EU data protection law. This section briefly accounts for those critiques, which mainly involved two elements of the notion of personal data: identifiability and its threshold, and the meaning of the syntagm ‘relating to’.

#### 6.2.1 Does Everything Relate to Everybody?

The A29WP’s opinion on the concept of personal data, and the assenting jurisprudence (i.e. the YS and Nowak cases),\footnote{CJEU, YS and others (n 845); Peter Nowak v Data Protection Commissioner (n 845).} clarified how the ‘relating to’ link between information and natural person can be constructed by the data’s content, purpose, or (likely) result. Both the A29WP\footnote{Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40).} and the CJEU interpreted ‘relating to’ in a broad manner, which gave rise to academic concern regarding the effects of such interpretation on the notion of personal data, and thus on the material scope of EU data protection law.

Purtova, for instance, claims that “in the age of the Internet of Things, datafication, advanced data analytics and data-driven decision-making, any information relates to a person in the sense of European data protection law”.\footnote{Purtova, ‘The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law’ (n 903) 3. Purtova’s article is, at the time of writing, the main scholarly critique to the current doctrinal and judicial construction of what ‘relating to’ means, and of its consequences for the tenability of the notion of personal data.} Granted, “some information is perceived as
relevant more easily’, but “when increasing amounts of data are gathered in real time from increasingly connected environments, intended to be used in automated decision-making about us, and we do not know how the autonomous self-learning and self-managing computers draw meaning from data, we should always reasonably assume that any information is likely to relate to a person, since we cannot eliminate this possibility with certainty”. Purtova argues that, in a “smart city where all aspects of the environment and people living in it are datified, and the inhabitants are subjected to a certain treatment in real time based on processing of the data, from the speed at which escalators are running to promote physical activity to the warmth and intensity of street lighting to prevent undesirable behaviour to targeted policing”, many categories of information which are not generally considered as ‘relating to’ natural persons will be increasingly linked with individual data subjects. “In such a world, any information within the ‘smart’ environment can be used and all information is likely to be used with the purpose of adapting the environment and impacting people”.

Playing the ‘devil’s advocate’, Purtova then highlights how a literal interpretation of the notion of personal data, and of the currently dominant doctrine and jurisprudence framing it, may render personal data also information traditionally not seen as personal, e.g. “rainfall per hour, temperature, wind direction and speed, together referred to as ‘weather’”, in the context of ‘smart cities’ and other large-scale ‘smart’ environments. She maintains that weather data is indeed information, and that – despite the fact that its content is not about a natural person – it may very well be used with a purpose involving (or likely to involve) a natural person. Even when the purpose of the weather data processing does not involve a natural person, she argues that its result is likely to make it relate to people. In such a ‘smart city’ context, Purtova assumes, weather data will thus relate to natural persons by virtue of its purpose or of its result.

It can be argued that such a reading can be overly broad, particularly if coupled with the low identifiability threshold discussed below. That line of reasoning could potentially lead to incongruous consequences even if one does not consider that, as Purtova does, “the ‘narratives of a frictionless world that surreptitiously adjusts the environment to the needs and
desires of its users’ are steadily on the way out of the realm of science fiction”. Such expansive interpretation of the ‘relating to’ link, and therefore of the notion of personal data, does not need to be applied to the data processed by a futuristic self-adapting ‘smart’ environment to make the material scope of the GDPR go haywire. Let us say that a person has a car insurance contract with an insurance company, and that the insurance premium varies according to the residence or domicile of the insured party. Such premium would be calculated according to a number of factors. Let us also say that one of them is the yearly average precipitations’ amount in the area – more rain might lead to more car accidents, and thus to a higher risk for the insurance company, that would factor that into its algorithm. Can we also claim that the yearly precipitations’ average rate of the area, as processed by the insurance company’s algorithm, is the customer’s personal data? It is information, the customer is a natural person who is readily identifiable, and the data can be used with the purpose of calculating the user’s premium, having the result of making him pay less or more. In light of the definition of personal data set by the GDPR, and of its interpretation as given by the A29WP, precipitation rates could thus be argued as being customers’ personal data. That line of reasoning is of course quite extreme, if not plainly absurd, but it can be sustained by relying on an extensive interpretation of the law and its doctrine and jurisprudence.

6.2.2 IS EVERYONE IDENTIFIABLE?

The natural person to which the information relates must also be either already identified or just identifiable for the definition of personal data to be integrated, and thus for the GDPR to be applicable. Many authors, from different disciplines, have however pointed out how the anonymity/identifiability dichotomy (on which the notion of personal data is based) is oftentimes blurry, warning that the line separating an anonymous person from an identifiable one is thinner than generally assumed. At the same time, the construction of the identifiability requirement in EU data protection law, doctrine, and jurisprudence, is quite expansive, so that the threshold after which to consider someone as identified or identifiable is quite low.

Any person – not just the controller or processor – can have the capability to identify a data subject. According to a strict interpretation of EU data protection law, controllers dealing with indirect identifiers must therefore consider, along with their own identification capabilities, the ones available to any other subject. If one considers, for instance, the possibility for the controller to contact the competent authority to identify people when necessary to initiate criminal proceedings, or the existence and operation of private investigation firms, it appears evident how the identifiability test adopted by EU data protection law and jurisprudence can potentially have a very low threshold.


1359 See supra (n 7, 11).


1361 As the Court does in CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845) §47.
The means that the controller or another person can theoretically use to identify the natural person to which the information relates are all the ones available at the time of processing, considering the state of the art, if likely reasonably to be used. In determining reasonable likelihood, account should be taken of all objective factors involving the processing at hand. As the CJEU stated in the Breyer case, the only instances where means are considered not likely reasonably to be used are the practical impossibility or the illegality of the means considered. EU data protection therefore sets, prima facie, an extremely low threshold for considering such natural person identifiable. In combination with the many ways in which information can be qualified as ‘relating to’ a natural person, this results in an expansive reading of the notion of personal data. At the same time, the faith held in the anonymity/identifiability dichotomy – and therefore on the notion of personal data – has been, to some extent, waning away.

The constant increase in the availability of auxiliary data, technological development, and know-how, contributed in rendering the border between identifiability and anonymity somehow fuzzy. Tracking individuals, offline and online, has proven to be a lucrative endeavour: an entire industry is focused on researching and developing ways to render people identifiable for a plethora of purposes, from behavioural advertising to cybersecurity to law enforcement. At the same time, computing became ubiquitous, thanks to smartphones and wearables, and everyday objects are increasingly being networked and digitalised. Physical spaces are being increasingly instrumented with sensors, in a merger between code and space – from the Internet of Things to the so-called ‘smart city’.

Anonymization – the process of irreversibly turning personal data into anonymous data – has also been under heavy scrutiny. Ohm’s stance provides an egregious summary of the concerns surrounding anonymization. Moving from the example given by three widely known re-identification attacks, Ohm worries that some powerful actors will eventually create a giant ‘database in the sky’, constantly feeding it with additional data until singling out specific individuals within ‘anonymised’ datasets becomes easy due to the amount of supplementary information available. It is essentially impossible to know the auxiliary

1362 GDPR, Recital 26.
1363 CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845) §46.
1364 ‘For years, it was widely believed that as long as data sets were ‘anonymized,’ they posed no risk to anyone’s privacy. [...] Unfortunately, the notion of perfect anonymization has been exposed as a myth’: Rubinstein and Hartzog (n 1296) 704.
1365 See Christl and Spiekermann (n 254); Christl, Kopp and Riechert (n 254). See also Zuboff, ‘Big Other: Surveillance Capitalism and the Prospects of an Information Civilization’ (n 1250).
1366 See Greenfield (n 108).
1367 See Kitchin and Dodge (n 10); Bratton (n 646).
1369 Ohm (n 42). After dealing with the purported failure of anonymization, and thus of identity protection, in a following contribution the author also muses about the future of attribute inference: see Ohm and Peppet (n 1254).
1370 Latanya Sweeney’s work, the Netflix re-identification attack (see Narayanan and Shmatikov (n 9)) and the AOL de-identification debacle (see Ohm (n 7)).
1371 This has been defined as the ‘Accretion Problem’; ‘once an adversary has linked two anonymized databases together, he can add the newly linked data to his collection of outside information and use it to help unlock other anonymized databases’: Ohm (n 42) 1746.
information an adversary has at its disposal, and thus threat modelling and the quantification of the risk of re-identification are intrinsically arbitrary. It has been also held that the expertise level needed to perform a re-identification attack, particularly when plenty of auxiliary information is available to the attacker, is not as sophisticated as to constitute a sizeable barrier anymore.

A dataset’s utility and the privacy it guarantees are moreover inherently at odds: significant gains in confidentiality (e.g. through obfuscation, generalization, or aggregation) are bound to diminish considerably the dataset’s granularity (and thus utility), and vice versa. In statistical databases, furthermore, it has been shown how it is impossible to formally guarantee that access to the database cannot enable an attacker to learn anything about an individual that could not be learned without access to the database, due to the attacker possibility to exploit auxiliary information. In any scenario where the attacker’s capabilities are not constrained by reference to an artificial threat model there will always be auxiliary information that, in conjunction with statistical data, allows to infer the identity of an unidentified subject, or some previously unidentified attributes of a known one.

Recital 26 of the GDPR, similarly to what Recital 26 of the Data Protection Directive (DPD), models the potential attacker as the data controller plus ‘another person’. It is not necessary that the means necessary to identify a data subject (e.g. auxiliary data contained in a private record) are directly available to the controller: they can be available to another person. Those means must also be ‘likely reasonably to be used’, which has been interpreted very broadly by the CJEU: only practical impossibility and illegality can rule out the means considered as available to the controller or to another person. That appears particularly far reaching if one considers that, in many cases, the ‘other persons’ whose means would have to be considered will be competent authorities, which often have the means to coerce information disclosure from third parties.

---

1372 Narayanan and Felten (n 1296).
1373 Narayanan and Felten (n 1296). Contra: ‘Linkage attacks, however, are much more complicated than they sound’ - Rubinstein and Hartzog (n 1296) 711.
1376 Simply put, attackers can draw from auxiliary data, extraneous from the anonymized dataset of reference; such auxiliary data, which cannot be modelled ex ante, can be cross-correlated with a purportedly anonymized dataset, potentially de-anonymizing the individuals to which the dataset refers: Dwork (n 1375) 2. Dwork proved that it is impossible, for a statistical database, to ensure that an attacker cannot learn, through the database, anything about an individual that could not be learned without access to such database: ‘in any ‘reasonable’ setting there is a piece of information that is in itself innocent, yet in conjunction with even a modified (noisy) version of the data yields a privacy breach’ - Dwork and Naor (n 1375) 93.
1377 GDPR, Recital 26; see also DPD, Recital 26. With respect to the DPD, Urgessa claims that ‘any other person’ means that a person is to be considered identifiable ‘if only another person than the controller is able to link a person to a data’ - Urgessa (n 1360) 522. The wording of Recital 26 of the GDPR, however, refers to ‘another person’, not to ‘any other person’ anymore: see s. 6.3.3.2 below.
1378 CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845) §46.
EU data protection law and jurisprudence adopt a very broad implicit adversarial model when framing indirect identifiability’s threshold, and thus consequently anonymity’s one. It has been underlined how, for the GDPR, almost anyone could be the adversary – the subject attempting the re-identification of a record, or the one possessing the auxiliary information necessary to that purpose. Data controllers must model their own identification capabilities assuming that they will be deemed as able to get to any auxiliary information available to any third party, unless practically impossible or illegal. A literal interpretation of EU data protection law and jurisprudence thus leans towards an absolute (or objective) stance on anonymity, where a person is considered identifiable by the controller if anyone can identify him or her, as opposed to a relative (or subjective) construction, where the person is considered identifiable only if the controller itself has the means to do so.

6.3 NARROWING DOWN PERSONAL DATA

The normative definition of personal data can thus be interpreted extremely broadly. It is, indeed, largely a matter of interpretation: the law in itself is dead letter, it is the reading given by the actor that applies it – the interpreter – that makes it ‘living’, fit to solve conflicts and to regulate behaviour. What will concretely set its scope is its interpretation by the qualified interpreter – the courts, lawyers, and all the actors involved in data processing endeavours. “(W)ords do not bind the interpreters; rather the interpreters give meaning to the words. [...] The critical people are the users, not the writers, of words.”

The interpreter is thus fundamental in defining the material scope of data protection law. Doctrinal and jurisprudential production can provide enough direction to guide the interpreter through an educated determination of what is personal and what is not. In this regard, despite adopting an apparently objective and formalist approach to

---

1379 With regard to identification, in the Breyer case (n 845), the CJEU adopted an objective (or absolute) criterion, where the natural person is deemed as identifiable if any subject can do so, and rejected the subjective (or relative) criterion, for which a person is deemed identifiable if the data controller can identify her by relying only on its own capacity. The Russian Federation, in contrast, reportedly adopted the subjective criterion: despite the fact that the personal data definition adopted by the Russian Federation corresponds to the one set by Convention 108, Russian judicial practice indicates that, when determining identifiability, courts consider only whether the data subject is identifiable by the controller alone: see Vladislav Arkhipov and Victor Naumov, ‘The Legal Definition of Personal Data in the Regulatory Environment of the Russian Federation: Between Formal Certainty and Technological Development’ (2016) 32 Computer Law & Security Review 868.

1380 See Urgessa (n 1360) 529.

1381 The CJEU ruled that the means to identify a data subject would not be considered as likely reasonably ‘if the identification of the data subject was prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant’: CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845) §46.

1382 See supra n 68.


1384 While this chapter’s jurisprudential analysis has been limited to the judgements of the CJEU, national courts and supervisory authorities provide additional, detailed guidance on what is personal data.

1385 See supra n 67.

1386 In the policy debate regarding the suitability of anonymization (and hence of the anonymity/identifiability dichotomy) as a basis for policy, Rubinstein and Hartzog identify two major groups: ‘formalists (for whom mathematical proof is the touchstone of any meaningful policy) and pragmatists (for whom workable solutions should prevail over theoretical concerns)’: Rubinstein and Hartzog (n 1296) 706.
identification, and a wide reading of the relational link, the A29WP also clarifies a few interpretative criteria to avoid the over-inflation of the concept of personal data. Likewise, the CJEU jurisprudence that can be read as opening up the door for an unchecked expansion of the material scope of EU data protection law does not have to be necessarily interpreted as extensively as recent literature suggests. While some elements of the doctrinal and judicial construction of the notion of personal data can indeed be framed as all-encompassing, the same literature and case law also contain guidance on how to keep them in check.

6.3.1 Flexibility by Design in the GDPR

The notion of personal data, to begin with, must be interpreted within its broader framework of reference – EU data protection law. This means that the finality of data protection must be taken into account when delimiting the scope of the notion of personal data. The GDPR lays down rules relating to the protection of natural persons with regard to the processing of personal data, and rules about the free movement of personal data. It protects fundamental rights and freedoms, and sets out that the free movement of personal data within the EU must not be restricted nor prohibited for reasons connected with data protection.

As the A29WP pointed out, the law’s subject matter and objectives play a substantive role in determining how to apply its provisions to a number of situations where the rights of individuals are not at risk. That appears to be even more true when considering the introduction of the so-called risk-based approach, which provides for additional flexibility in the law’s application.

Even when data is personal, the applicability and application of the GDPR are conditioned by several exemptions and derogations. Aside from what is tout court excluded from the GDPR’s material and territorial scopes, EU data protection embeds a degree of flexibility that mitigates the width of the definition of personal data it adopts, or the consequences thereof. The breadth of the notion of personal data should not automatically lead to overstretching the scope of the GDPR, nor to the application of its rules to situations which

---

1387 E.g. inter alios Aldhouse (n 1296); Purtova, ‘The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law’ (n 903).


1389 I.e. the protection of fundamental right and freedoms with regard to the processing of personal data, and the free movement of such data.

1390 See YS and others v. Minister of Immigration, Integration and Asylum (n 845) §41-42.

1391 GDPR, Art. 1.

1392 On the other hand, the Working Party also cautions against an overly restrictive interpretation, which would leave individuals deprived of protection. ‘The scope of the data protection rules should not be overstretched’: Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40) 5.


1394 GDPR, Art. 2, 3.

1395 E.g. GDPR, Art. 23; Art. 85 ss.
were not meant to be regulated through data protection law. At the same time, an overly restrictive interpretation of the concept should be avoided, in that it would deprive individuals from the protection that the legislator meant them to receive.

Apart from the holistic approach to the GDPR, and the teleological reading of the concept of personal data in light of the function of the right to data protection, there are additional factors that can guide the interpreter in determining the extent of the notion of personal data in a reasonable manner. Data protection law does not apply to personal data in a vacuum, but to its processing. That is to say that each concrete processing instance has additional context that is bound to be considered when determining whether the data is personal. Each distinct case has additional contextual elements that, read through the interpretative canons provided by law, doctrine, and jurisprudence, (co)determine the information’s personal (or anonymous) nature.

European data protection law is a highly convoluted framework, cutting through different normative levels, and operating across many jurisdictions. While legal certainty is undoubtedly an objective to pursue, the complexity of the normative framework of reference derives from the intricacy of the matter at hand: balancing the right to data protection with the free movement of data. The contextual and relative character of the definition of personal data, while possibly detrimental to legal certainty, allows European data protection to be flexible and technology-neutral enough not to leave any data processing situation that can have an impact on individuals outside of its scope. There is, in other words, a natural and unavoidable trade-off between the flexibility of data protection law and the legal certainty that a narrower notion of personal data would grant.

6.3.2 The interaction between identifiability and the ‘relating to’ link

Any information – regardless of its content – can be used for a purpose involving an individual, or can be processed in a way that results in an impact for a natural person. Any kind of data can possibly relate to people. That does not, however, suffice to make the information personal data on its own: the natural person to which it refers must also be identified or identifiable. I argue that, in situations where the ‘relating to’ link is integrated by the purpose or result elements, its interaction with the identifiability requirement functions as a logical limit.

---

1397 E.g. the subjective, restrictive interpretation adopted by the courts in the Russian Federation: see Arkhipov and Naumov (n 1379).
1398 See Commission of the European Communities, ‘Commission Communication on the protection of Individuals In relation to the processing of personal data In the Community and Information security’ COM(90) 314 final. Brussels, 13.09.1990, p. 19: ‘As in Convention 108, a broad definition is adopted in order to cover all Information which may be linked to an individual. Depending on the use to which it is put, any Item of data relating to an individual, harmless though It may seem, may be sensitive […] to avoid a situation in which means of indirect identification make it possible to circumvent this definition’.
1399 In other words, wondering ‘is this personal data?’ is a disingenuous exercise, in that it does not lead to the same reasoning as ‘is this the processing of personal data?’ by removing the contextual and environmental considerations that, when data does not relate to a person by virtue of its content, make or break its qualification as personal.
to the potential over-inflation of the concept of personal data, and thus of the material scope of EU data protection law.\(^{1401}\)

When the relational link between information and natural person is justified through purpose or result, rather than through content, additional information will necessarily be required to render such person identified or identifiable, and thus the data personal. When the information relates to the data subject due to its content, the very substance of the information may lead to the identifiability of the data subject – its content is about that specific person in that it has biographical significance. An employee ID identifies through its content, no matter how it will be used, or the result its processing will have. When the relational link is given by the result or the purpose one, however, additional information will \textit{always} be necessitated to satisfy the identifiability requirement, and hence the definition of personal data. If the data could lead to the identifiability of the data subject on its own, the relational link would be integrated by the content element, without the need to recur to the purpose or result ones. Let us consider a situation along the lines of the example above,\(^{1402}\) where information generally thought as anonymous in the sense of ‘apersonal’\(^ {1403}\) (e.g. average precipitation rates in a given area) can possibly become personal by virtue of its purpose or of the result of its processing (e.g. for the calculation of an insurance premium): that can only happen in conjunction with auxiliary information leading to the identification of the data subject considered.

In a way, qualifying the data as relating to a natural person through the purpose or result element naturally leads to a higher identifiability threshold than the one required when the information relates to the data subject through its content. To be clear, the relationship test and the identifiability test are undoubtedly different assessments, aiming at evaluating two distinct components of the notion of personal data. What I argue is that the ‘relating to’ and the identifiability elements of the concept of personal data are logically related, tied together, so that the potential expansive effect of the ‘relating to’ element of notion of personal data as framed by the Article 29 Working Party\(^ {1404}\) and by the CJEU in the Nowak case is reined in through the identifiability test. If the \textit{content} of a given set of data does not relate to the data subject, and the link must be found in its \textit{purpose} or \textit{result}, the identification of the data subject will necessarily have to happen through auxiliary information: the data is not \textit{about} the natural person in itself\(^ {1405}\). Since direct identifiers relate to the data subject due to their content, there

\(^{1401}\) Methodologically, the analysis has been carried out by enucleating all the possible configurations of the interaction between the ‘relating to’ link and the identifiability requirement (i.e. when the data subject is identified or identifiable, directly or indirectly, through the means available to the controller or to another person, and the data relate to him or her by virtue of their content, or their purpose, or their result), systematising them in a table, and populating the table with examples of processing instances strictly suiting each possible configuration (e.g. information relating, by virtue of its purpose, to a natural person that is indirectly identifiable through the means available to a person other than the controller, or information relating, by virtue of its content, to a person that is directly identifiable by the controller).

\(^{1402}\) As opposed to information that is anonymous by virtue of the fact that it went through an anonymisation process.

\(^{1403}\) Admittedly, what ‘information relating to a natural person by virtue of its content’ means is still largely up for debate. I take a somewhat narrow view, and consider it as data whose content relates to the data subject no matter how it is used or the result of its processing, e.g. an employee number. An IP address is information whose content is about a machine, but whose purpose or result may relate to the natural person using that machine. Conversely, my employee ID number is information whose content relates to me no matter its purpose


\(^{1405}\)
is never the need to qualify them as relating to the data subject due to their purpose or result – the elements that justify the relational link are alternative, not cumulative. Data deemed as relating to a natural person due to its purpose or result, but not to its content, needs however to be tied to additional auxiliary information that render the person identified or identifiable to become personal.

In other words, the information must be actually relating to an identifiable data subject, not merely relatable. The wording of the definition of personal data is clear: while the person can be just identifiable, the information must be necessarily related, not just relatable. The legislator explicitly avoided referring to the mere possibility of a relation, requiring an actual link between the information and the identified or identifiable natural person. If the link is satisfied through the content element, then of course the information and the natural person will be related, not just relatable – the information is about the person in the most literal sense. Even when the relational link is integrated by the purpose or result elements, the data will still have to be actually related to a specific person, rather than just relatable. That must necessarily happen through additional auxiliary information that makes the data subject individually distinct from the group of individuals to which the information is just relatable.

Data protection is an individual right, its collective dimension currently marginal. It is meant to protect individuals, rather than the groups to which they belong. It does not protect all subjects to which an attribute (e.g. affluence, or being a single mother) pertains, as a collectivity, but does shield them from the moment where they become personally identifiable. Therefore, even if any information can possibly relate to natural persons due to its purpose or its result, despite the fact that its content does not relate to a natural person by itself, the data always needs to be actually tied to auxiliary information connecting it with the data subject interested by that concrete processing instance to become personal. Data about things may very well become personal data, but only when it can be tied through auxiliary data to an identified or identifiable natural person in a specific processing instance.

The interaction between the ‘relating to’ link and the identifiability requirement makes it so that the configuration of the ‘relating to’ nexus as satisfied by virtue of the purpose or result or impact. In this sense, I believe the CJEU, in its Nowak judgement, framed the concept of ‘relating to through content’ in a confusing manner, conflating the content element and the purpose and result ones: see e.g. ‘the content of those answers reflects the extent of the candidate’s knowledge and competence in a given field and, in some cases, his intellect, thought processes, and judgment’ (Nowak (n 845) §37) and ‘(t)he content of those comments reflects the opinion or the assessment of the examiner of the individual performance of the candidate in the examination, particularly of his or her knowledge and competences in the field concerned’ (Nowak (n 845) §43).

1406 An anonymous reviewer suggested a helpful parallelism with Durant v Financial Services Authority [2003], EWCA Civ 1746. In Durant, the Court of Appeal of England and Wales tied the qualification of information as personal data to its existence in ‘a continuum of relevance or proximity to the data subject’, to be determined considering ‘whether the information is biographical in a significant sense, [...] going beyond the recording of the putative data subject’s involvement in a matter or an event that has no personal connotations’ and that the information ‘should have the putative data subject as its focus’ (§28). The Durant v FSA criterion of biographical significance is arguably too narrow and in contradiction with successive ECJ jurisprudence. However, the judgement still provides helpful guidance on the interaction between identifiability and the ‘relating to’ link when it recognizes as a ‘reason for hesitation [...] that in some cases it is Mr. Durant’s identity that leads to the information, rather than the information leading to Mr. Durant’ (§78).

1407 See Mantelero, ‘Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection’ (n 343); Alessandro Mantelero, ‘From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era’ in Linnet Taylor, Luciano Floridi and Bart van der Sloot (eds), Group Privacy (Springer 2017).
element, rather than by the content one, leads to the necessity of a higher amount of auxiliary information to integrate the notion of personal data. In a way, it functions as a system to differentiate data about people from data about a person, and hence data which is encompassed by the material scope of EU data protection law from information that is not. Granted, however, data about people can become data about a person, if so processed: the qualification of data as personal (or, conversely, anonymous) must be performed considering the information within its lifecycle, rather than statically.

6.3.2.1 DATA LIFECYCLE

If any information can somehow relate to natural persons, and identifying someone is gradually becoming easier and easier, then all data can potentially be personal, and data protection’s scope becomes gargantuan. Such concern is legitimate, and a valid argument to make when questioning the regulatory structure and mechanics on which EU data protection is based; less so when applied to the concrete determination of what personal data within an actual processing instance is. Even if it is true that all data can become personal data, that does not mean that such data is personal all the time. Information relating to a natural person by virtue of the purpose or result elements will not necessarily be personal throughout its entire lifecycle.

Data has indeed a lifecycle: 1408 it gets created, collected, processed, re-shaped, aggregated, stored, and eventually deleted. When it links to a natural person through its content, it will likely do so throughout its lifecycle, unless the content changes. When the information is linked to the natural person only through the purpose or result elements, it will relate to them just for a specific segment of its lifecycle. A passport number, data whose content relates to an individual, will keep relating to that individual throughout its entire lifecycle. Data about the performance of a vehicle will relate to an individual only if and when it has a purpose or result actually involving that particular person, for instance when it is used to evaluate a driver’s performance. Likewise, let us assume that a public administration shares information with the public, for instance public transport information, which then gets used by another party (e.g. an employer) to generate particular consequences involving an identifiable individual (perhaps to see whether an employee’s unpunctuality can be blamed on public transport). Such data could be personal, in connection with an individual’s identifiers, only from the moment when it is used to generate an effect involving such individual (i.e. when it relates to him through the result element), since nor the information’s content nor its purpose relate to natural persons.

In the processing instances where the relational link is justified through the purpose or the result element, rather than through the content one, the information considered will not necessarily 1409 qualify as personal from its inception to its erasure, but only if and to the extent to which it is meant (or likely) to be used to generate effects involving specific natural persons. Moreover, as highlighted in the previous section, it will still need to be processed in connection with additional information that can lead to their identification to be qualifiable as personal. I believe it would be preferable to shift the general focus of the debate around the inflation of the material scope of data protection law, from the static notion of personal data, to the


1409 Information that does not relate to a data subject due to its content but by virtue of its purpose or result can still happen to relate to the data subject for the entirety of its lifecycle, depending on the kind of data and how it is processed, but that is a mere possibility. Information whose content relates to a data subject will keep relating for a data subject for its entire lifecycle no matter how it is processed.
processing of personal data. Data protection law does not indeed apply to personal data in itself, but to its processing. As a naturally relational concept, personal data is better understood when considered, dynamically, within the concrete processing instance evaluated. The purported expansive effects deriving from the combination between the low identifiability threshold and the wide range of ways information can relate to a natural person, as set by EU data protection law and jurisprudence, seem to be much less substantial when considering personal data that relate to the data subject by virtue of their purpose or result within their lifecycle, rather than statically.

6.3.2.2 ATTRIBUTE PROTECTION

The core purpose of data protection is safeguarding individuals from the harms that might derive from unfair information processing. The notion of personal data, on which EU data protection is based, is meant to be potentially applicable to any kind of data, since any information can be hypothetically used to generate harmful consequences affecting human beings. In other words, data protection law does not aim at protecting individuals only from the misuse of their identities, but from the misuse of their attributes – their characteristics and defining traits – too. From this perspective, the fact that any data can become personal data is a regulatory ‘feature’, rather than a ‘bug’, necessary to provide the flexible, contextual protection in context EU data protection is meant to afford. The concept of personal data must be framed diachronically: the exact same piece of information can be anonymous or personal depending on the context, actors, and time of processing. Attributes, when the natural person they are linked to is not identifiable, are not directly covered by EU data protection legislation: they are data about people, rather than data about a person. That remains true up to a certain point in the information’s lifecycle – the application of the inference to a specific natural person – after which data protection law kicks in.

Statistical information – e.g. an anonymised, aggregated dataset – for instance does not allow the singling out of specific individuals from the sample of reference. Its proper sanitization prevents identity inference, impeding that an attacker identifies a specific individual within the record. Depending on the concrete circumstances of the case, the attacker can however often infer meaningful attributes about a data subject, without being able to identify the record linked with that specific individual. EU data protection law, as an individual right, does not safeguard from group inferences per se, but does grant attribute protection as soon as the processing purpose or result relate to a specific natural person, without limiting itself to identity protection. It does so by considering personal data also as information whose content does not directly relate to natural persons, but that is still tied, through its purpose or result, to specific individuals through auxiliary information that render them identifiable. For instance, aggregated data showing the average income in my neighbourhood cannot be personal on its

---

1410 Along with ensuring the free movement of personal data: GDPR, Art. 1.
1411 ‘What matters here, however, is not simply whether the individual with prior knowledge can identify the data subject concerned but whether he/she will learn something new from the information obtained through re-identification’: Article 29 Data Protection Working Party, ‘Opinion 06/2013 on Open Data and Public Sector Information (‘PSI’) Reuse WP207’ (n 521) 15.
1412 See e.g. Schwartz and Solove (n 1400) 877; 892.
1413 Granted, the same information can relate to more than a single natural person; it may also constitute personal data for each of them, provided that each person is also identified or identifiable (see CJEU, Nowak (n 845) §45). See also Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP 136’ (n 7) 11: ‘the same piece of information may relate to different individuals at the same time, depending on what element is present with regard to each one’.
own: I cannot be singled out from the aggregated record, and statistical information is merely relatable to me, not actually related. As soon as the very same information gets linked with additional identifiers to be used with a particular purpose or result involving me as a specific identifiable individual – for instance, if it gets used as a proxy for credit rating – such data does however become personal.1414

Any data can theoretically generate informational harm: not just the information whose content refers to an individual, but whichever data can be used with a particular purpose or to generate relevant consequences for the data subject. As the Article 29 Working Party states,1415 the capacity to infer information about a person (i.e. one of her attributes) must be considered when assessing the level of protection anonymization should afford, along with the possibility to single her out or to link different records concerning that person.1416 The mere possibility to infer an attribute applicable to a group (e.g. the average income in a certain area code) does not however render the data personal by itself: it merely makes it relatable to a person. The inference becomes personal when it begins to be actually related to the data subject, which requires auxiliary information leading to that specific person’s direct or indirect identifiability. EU data protection is thus meant to safeguard whichever attribute might characterise a person, but only from the moment where the group inference starts to be related to a specific identifiable person.1417

6.3.3 MODULATING IDENTIFIABILITY

The identifiability test adopted by European data protection law and jurisprudence is, potentially, very broad. Natural persons are deemed identifiable if the controller or another person can single them out, directly or indirectly, through all means likely reasonably to be used – i.e. unless those means are practically impossible or illegal1418 – in consideration of the state of the art. Individuals can theoretically be identified through several means and by many actors. At the same time, singling people out from a group has arguably never been easier for a motivated attacker. Competent authorities, for instance, are well within the meaning of ‘another person’, and have often the power to compel the production of the auxiliary information necessary to tie all sort of data to specific natural persons. An order from a competent authority will thus often be a ‘means likely reasonably’ to be used to identify the natural person to which the information relates. This logically leads to a low threshold after which a person is to be considered as legally identifiable.

1414 Borgesius follows the same line of reasoning with regard to online behavioural advertising: ‘Some data processing activities for behavioural targeting do not concern personal data. A company can use personal data to construct a model […] Such models do not consist of personal data, as they do not relate to a specific person. As soon as a company applies the model to an individual, however, the information relates to this person because of its purpose or its result’ - Frederik J Zuiderveen Borgesius, ‘Singling out People without Knowing Their Names–Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation’ (2016) 32 Computer Law & Security Review 260.
1415 Article 29 Data Protection Working Party, ‘Opinion 05/2014 on Anonymisation Techniques WP216’ (n 521) 3. On the opinion, see El Emam and Alvarez (n 1109).
1416 ‘Identification’ not only means the possibility of retrieving a person’s name and/or address, but also includes potential identifiability by singling out, linkability and inference’: Article 29 Data Protection Working Party, ‘Opinion 05/2014 on Anonymisation Techniques WP216’ (n 521) 10.
1417 See Zuiderveen Borgesius (n 1414) 260.
1418 CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845) §46.
This should not however lead an “identifiability assumption”, where – in case of doubt – individuals are to be considered as identifiable by default under the GDPR. While the necessity to avoid elusion mandates the adoption of a low and flexible identifiability thresholds, EU data protection law, jurisprudence, and doctrine, also provide a number of criteria and considerations through which to modulate the identifiability test.

6.3.3.1 MEANS TO IDENTIFY AND REASONABLE LIKELIHOOD

Determining identifiability must be done on a case-by-case basis, for each concrete processing instance, through the interpretative criteria made available by the law and its jurisprudence. The mere theoretical possibility of identification is not sufficient to render someone identifiable: such possibility must be concrete, modelled according all objective factors of the processing.\textsuperscript{1419} The ‘means likely reasonably to be used by the data controller or by another person’ should be interpreted in light of the functioning of EU data protection law, which singles out each processing instance by reference to the data processed, the actors involved, and the purposes of the processing.

To ascertain whether means are reasonably likely to be used to identify the natural person, one should consider all objective factors pertaining to the processing:\textsuperscript{1420} account must be taken of ‘all’ the means ‘likely reasonably’ to be used for identification by the controller and third parties, paying special attention to the current state of technology, and the constant increase in computing power, know-how, and tools available.\textsuperscript{1421} However, mere theoretical chance is not sufficient to deem a person identifiable; if, taking into account “all the means likely reasonably to be used by either the controller or another person”, that possibility does not exist or is negligible, the person should not be considered as identifiable.\textsuperscript{1422} That is the case if the identification of the data subject is prohibited by law, or practically impossible, which happens when the risk of identification is deemed, in reality, to be insignificant.\textsuperscript{1423} ‘Objective’ factors, I believe, should be interpreted as referring to the contextual and environmental elements of the processing, rather than to the mere existence of the means that can be used to identify a person.

To be clear, I am not arguing for a switch to a paradigm where identifiability depends on the subjective, relative capacity of the entity performing the identification attempt – the controller or ‘another person’. I believe, however, that artificially removing contextual considerations from the identifiability assessment can lead to perverse results, as much as basing the assessment on the capabilities of the controller alone would render data protection law too easily avoidable. The objectivity of the factors to be considered when determining identifiability should be referred to the environment in which the controller operates: the CJEU

\textsuperscript{1419} GDPR, Recital 26; Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40) 15 ss.
\textsuperscript{1420} GDPR, Recital 26.
\textsuperscript{1423} CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845) §45-46. It is reasonable to assume that the Court did not mean to equate impossibility with zero probability, but with a negligible level of the latter – ‘so that the risk of identification appears in reality to be insignificant’.

does indeed refer to *practical* impossibility,\textsuperscript{1424} not to impossibility *tout court*, when discussing
the means that are not likely reasonably to be used to identify the data subject.

The GDPR exemplifies the factors to be considered when determining a person’s identifiability
by referring to the costs of and the amount of time required for identification, taking into
consideration the state of the art.\textsuperscript{1425} Other important factors are, for instance, the intended
purpose of the processing, the way it is structured, the advantage expected by the controller
and the interests at stake for individuals, as well as the risk of organisational and technical
failures.\textsuperscript{1426} Indeed, the subjective motivation of the attacker might impact the identifiability
threshold: some datasets will be more desirable than others. For example, information might
be more likely to lead to the identification of the data subject if it has a significant commercial
value, or if it can be used for law enforcement or intelligence purposes. Attackers might be
also drawn if the information reveals newsworthy information about public figures, or if it can
be used for political purposes, or even if it could just raise curiosity.\textsuperscript{1427}

The availability of auxiliary data that might be used to identify the data subject must be
contemplated too, both in terms of its public availability as well as considering how certain
attackers might have access to privileged information that allow to identify individuals.
Identification risks increase where someone is likely to know a large amount of auxiliary data
about a person, such as family members, colleagues, doctors, or other professionals. The
auxiliary data needed to perform re-identification could of course also be information available
to businesses or organisations, including law enforcement, judicial authorities, or other public-sector bodies; it could also be published on the Internet, available to everyone.\textsuperscript{1428}

The identifiability test is meant to be dynamic, and should consider both the state of the art at
the time of the processing and the possibilities for future development during the foreseen
processing period. Storage time is a paramount factor in determining identifiability’s threshold.
What may not be likely reasonably possible today, or in a month, might become feasible in a
decade. Identification must be put in relation to the information’s lifecycle, and thus the
controller should consider the possibility of future identification, which may make the data
personal (from that moment on, not retroactively). Data controllers should stay aware of
relevant developments, and enact the necessary technical and organisational measures.

Technical and organisational considerations are particularly important in determining the
factors that may render the data subject identifiable, and thus the controller might consider
testing the likelihood that the identification, if attempted, would be successful.\textsuperscript{1429} In some
circumstances it can be difficult to establish such risk, particularly where complex statistical
methods might be used by a third party to match various pieces of anonymised data. It can be
good practice to perform a re-identification test to ascertain the potential for re-identification,
attempting to re-identify individuals from the datasets that are being assessed. Such “motivated intruder” test essentially involves considering whether an attacker would be capable to achieve re-identification, if adequately motivated. The approach assumes that the attacker is motivated, competent, and has access to resources commensurate with the motivation it may have for the re-identification.

The criteria that have been provided to help in the interpretation of the notion of personal data, particularly with respect to the identifiability test, seem to mirror what has been defined, in the field of statistical confidentiality, as the ‘data environment’: the set of formal and informal structures, processes, mechanisms and agents that act on, define, control, or interact with data, or provide interpretable context for such data. Each data environment has been framed as consisting of four components: data, agency, governance, and infrastructure.

The data element considers the information available within the environment; the agency one models how agents might act on and in the environment; the governance element identifies the discipline of the users’ interaction with the data; and the infrastructure component considers the physical and logical structures and processes that regulate the data environment. While the concept of data environment pertains to the statistical confidentiality milieu, it might well provide for a way to formalise identifiability – and perhaps the status of information as personal tout court – within the context of the right to personal data protection.

6.3.3.2 From ‘Any other person’ to ‘Another person’

The interpretation of what constitutes a means reasonably likely to identify a person is not the only factor that can lead to the inflation of the identifiability component of the notion of personal data. The fact that identifiability must be tested against the means available to both the controller and ‘another person’ can also lead to an excessively low identification threshold, and hence, in conjunction with the width of the ‘relating to’ link, to the disproportionate width of the definition of personal data. In this regard, I hold that the ‘another person’ Recital 26 of the GDPR refers to should be interpreted teleologically: it should be ‘likely reasonably’ for the controller or the processor to have access to the third party possessing the auxiliary information that permit the identification of the data subject.

---

1430 Article 29 Data Protection Working Party, ‘Opinion 06/2013 on Open Data and Public Sector Information (‘PSI’) Reuse WP207’ (n 521) 17.
1432 Article 29 Data Protection Working Party, ‘Opinion 06/2013 on Open Data and Public Sector Information (‘PSI’) Reuse WP207’ (n 521) 16.
1433 See Elaine Mackey and Mark Elliot, ‘Understanding the Data Environment’ (2013) 20 XRDS: Crossroads 36; Elliot and others (n 1374).
1436 In this sense, see Opinion of AG Campos Sánchez-Bordona in case Breyer v Germany, C-582/14, 12 May 2016, ECLI:EU:C:2016:339, 67-68.
1437 See the A29WP on how the disclosure of public-sector information impacts the assessment of the ‘other person’ who may re-identify the data subject: ‘once data are publicly released for reuse, there will be no control over who can access to the data. The likelihood that ‘any other person’ will have the means and will use those means to re-identify the data subjects will increase very significantly. Therefore, and irrespective of the
party with powerful means to which the controller cannot actually have access should not be considered as ‘likely reasonably’ to be used to identify a person.

To determine whether a natural person is identifiable, the GDPR mandates the consideration of all the means reasonably likely to be used, either by the controller or by another person, to identify the data subject, directly or indirectly. ¹⁴³⁸ For data to be personal, it is not necessary that it alone identifies the data subject; it is not required for all the information enabling the identification of the data subject to be in the hands of a single entity, either ¹⁴³⁹ – as long as access to that entity is reasonably likely for the data controller. Indeed, a strict literal interpretation of the concept of personal data could expand its scope to all information all the time, regardless of the information’s inability to reveal the data subject on its own: “(i)t would never be possible to rule out, with absolute certainty, the possibility that there is no third party in possession of additional data which may be combined with that information and are, therefore, capable of revealing a person’s identity”. ¹⁴⁴⁰

Just as the means that may be used by the controller must be constrained to the likely reasonably ones only, so the third parties who may be approached by a controller to identify the data subject should be understood as the ones that can likely reasonably be accosted. Reasonable likelihood would not occur when contact with those third parties is exceedingly costly, considering both human and economic capital, practically impossible, or prohibited by law. Such an interpretation has been put forth by Advocate General Campos Sánchez-Bordona in the Breyer v Germany case; ¹⁴⁴¹ while the CJEU did not reject it, it did not explicitly confirm its validity either. I believe clarifying that the same ‘reasonable likelihood’ test applicable to the means for identification could (and should) be applied to the ‘other person’ that can identify the data subject in lieu of the data controller would have provided valuable guidance from the CJEU.

In this sense, it seems significative to point out a notable difference between Recital 26 of the GDPR, and its predecessor, Recital 26 of the DPD. While Recital 26 of the DPD specified that, to determine whether a person is identifiable, one should consider “all the means likely reasonably to be used either by the controller or by any other person”, Recital 26 of the GDPR indicates that “account should be taken of all the means reasonably likely to be used [...] either by the controller or by another person”. That would seem a meaningful semantic difference. The syntagms ‘any other person’ and ‘another person’ are not equivalent: the former conveys that the means to be considered when determining the identifiability of a data subject can be available to any person; the latter clarifies that such means do not necessarily have to be available to the controller, but can be available to another person too. It is still too early to see how courts will interpret this lexical change, but moving from the postulation that the concept of personal data in the GDPR mirrors the one in the DPD, it is reasonable to assume that the legislature meant to clarify that there are situations where the fact that a person is identifiable interpretation of recital 26 in other contexts, when it comes to making data available for reuse [...] utmost care should be taken to ensure that the datasets to be disclosed should not include data that can be re-identified by means likely reasonably to be used by any person, including potential re-users, but also other parties that may have an interest in obtaining the data, including law enforcement’ - Article 29 Data Protection Working Party, ‘Opinion 06/2013 on Open Data and Public Sector Information (‘PSI’) Reuse WP207’ (n 521) 13.

¹⁴³⁸ GDPR, Recital 26.
¹⁴³⁹ CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845) §41-44.
¹⁴⁴⁰ Opinion of AG Campos Sánchez-Bordona in case Breyer v Germany (n 1436) §65.
¹⁴⁴¹ Opinion of AG Campos Sánchez-Bordona in case Breyer v Germany (n 1436) §68.
by a particularly resourceful controller does not make that person identifiable by any controller by default.

The breadth of the identifiability element can thus be tempered by a number of considerations and parameters which, while guiding in the interpretation of the concept of personal data, also contribute in narrowing it down to a reasonable dimension. The flexibility and contextual nature of the notion of personal data requires a case-by-case approach to determining whether information is personal or not. The discretion of the interpreter is however bound by a number of factors that, if followed, would prevent outcomes deviating from the ratio legis of EU data protection law.

6.4 OPEN DATA, DATA PROTECTION, AND AUXILIARY INFORMATION

The previous chapter highlighted the incompatibility between open data and data protection, while the preceding section pointed out some of the interpretative criteria that can be used to avoid an overexpansion of the scope of the concept of personal data, and hence of the material scope of data protection law. In a way, they dealt with the impact that personal data protection has upon open data sharing and re-use. Conversely, it is generally well understood that open data is meant not to be prejudicial to personal data protection, and that personal data are in practice never shared as open data. Yet, open data publishing and re-use, even if focused on anonymous information, can still have an impact on individuals’ data protection rights.

There are mainly three ways in which open data can compress individuals’ rights to privacy and personal data protection: (1) through the inadvertent breach of the applicable law, (2) by allowing the inference of sensitive attributes about the population of reference without triggering the applicability of data protection legislation, and (3) by augmenting the pool of auxiliary information available to an attacker that means to identify a particular person from an anonymized (or de-identified) dataset. The first impact is obviously not specific to open data, and is almost self-explanatory: sometimes, be it for ignorance or negligence, some act within the open data ecosystem may breach data protection inadvertently — for instance by not considering MAC or IP addresses as personal data due to a lack of domain knowledge.

The second kind of impact relates to the possibility of drawing inferences from an open dataset about some attributes pertaining to a particular population sample — for instance, residents of a particular area, smokers, or single mothers — even when such dataset is anonymous or has been anonymised. Open data does not need to refer to a specific natural person to engender negative consequences: in other words, anonymization does not address the possibility to

---

1442 I.e. information obtained outside the system of reference (e.g. a single open data set, or all the datasets in an open data portal) that can be cross-correlated or otherwise processed with the information.

1443 The semantics here are interesting — defining a dataset as anonymous seems to assume the impossibility of re-identifying the data subjects within it, which has been proven impossible, at least at a theoretical level. Defining it as de-identified seems more fitting, as it does not rule out the possibility of re-identification, and yet that is not the vocabulary adopted by EU data protection law, where what is not personal is anonymous - see Recital 26 of the GDPR; Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40); Article 29 Data Protection Working Party, ‘Opinion 05/2014 on Anonymisation Techniques WP216’ (n 521).

1444 The inadvertent breach of applicable legislation is, incidentally, one of the risks deriving from open data sharing and re-use that has been highlighted in literature: see e.g. Kucera and Chlapek (n 1169); Zuiderwijk and Janssen (n 620).

1445 See CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845).
misuse the data, for instance by building a discriminatory model to be used as a basis for automated (or semi-automated, for that matter) decision-making.

Access controls and use restrictions are essentially incompatible with the concept of open data, and thus controlling how an open dataset is used after its publication is a fundamentally wicked problem. Integrating the material scope of EU data protection law requires a data subject to be identified or identifiable, while attribute inferences are group inferences, so this kind of processing does not necessarily trigger its applicability, at least until the inference gets used with a purpose involving an identifiable natural person, or until the result of the processing generates some kind of consequences for an identifiable natural person. The third impact that open data sharing and re-use can have on personal data protection derives from the fact that the publication of data increases the size of the pool of information available to an attacker that aims at re-identifying a data subject from an anonymized record, or to draw inferences about some (sensitive) attributes of a known data subject. Again, this is not specific to open data per se, but it is amplified by the heightened availability, accessibility, and re-usability of open data as compared to other modalities of information sharing.

The third impact that open data sharing and re-use can have on personal data protection derives from the fact that the publication of data increases the size of the pool of information available to an attacker that aims at re-identifying a data subject from an anonymized record, or to draw inferences about some (sensitive) attributes of a known data subject. Again, this is not specific to open data per se, but it is amplified by the heightened availability, accessibility, and re-usability of open data as compared to other modalities of information sharing.

The issue is that it is impossible to formally guarantee that an attacker does not obtain from an anonymised record information that he could not obtain without access to that record, as it is not possible to know what kind of auxiliary data an actual attacker would have at its disposal when trying to infer knowledge from a purportedly anonymous dataset. Information shared as open data is likely to be available, as it is freely accessible and re-usable by anyone for any purpose, and in a format that maximises its interoperability, and thus the possibility to analyse it. A simplistic example might better illustrate the mechanics: the information that a pseudonymous individual earns 75% more than the average within a given area does not reveal how much that individual earns unless one knows how much that area’s average is. The issue is far from being merely theoretical: there are many examples of re-identification attacks performed on a dataset by using various kinds of auxiliary information available to the attacker to single out individual records. In addition, the amount of information generated by and about individuals is ever-increasing, and the know-how required to perform a re-identification attack is becoming more and more commonplace.

Open data, as (and more than) any other kind of data release, thus increases the pool of auxiliary data that an attacker can use to identify a subject within an anonymous record, or to infer an unknown attribute about a known subject. That does not however mean that

---

1446 For instance, statistics about teen pregnancies in a particular geographic area are not personal data per se, but may very well become personal if they are used for the credit scoring of specific individuals.

1447 A nosy neighbour in a rural, unpopulated area might dig into a dataset reporting on farm subsidies to figure out how well the proprietors of the farm down the road are doing. The same might go for a real estate company trying to gauge how motivated are the owners of a land plot it wants to buy before making an initial offer.

1448 See Dwork (n 1375); Dwork and Naor (n 1375). See also Ohm (n 42); Wu (n 1296); Rubinstein and Hartzog (n 1296).

anonymizing a dataset is never possible anymore,\(^{1450}\) nor that it is now trivial to infer new attributes from other known attributes, so that “everything reveals everything”.\(^{1451}\) It just means that some formal definitions of privacy in a narrow sense (privacy as non-disclosure, confidentiality) are impossible to guarantee the same way semantic security is guaranteed in cryptosystems.\(^{1452}\)

The issue has been clarified by Dwork\(^{1453}\) when making the case for differential privacy – a formal mathematical framework for quantifying and managing privacy risks;\(^{1454}\) I will refer to her explanation. Technical conceptions of privacy, for instance in the context of statistical data disclosures, require formal definitions.\(^{1455}\) In 1977, Tore Dalenius\(^{1456}\) defined the privacy requirements that a statistical disclosure \(S\) should comply with (Dalenius’ desideratum) by stating that, for an object \(O_k\) that has a value (or attribute) \(D_k\), “(i)f the release of the statistics \(S\) makes it possible to determine the value \(D_k\) more accurately than it is possible without access to \(S\), a disclosure has taken place; more exactly, a \(D\)-disclosure has taken place”.\(^{1457}\) In other words, according to Dalenius’ formal definition of privacy, it should not be possible to figure out something from the data that it cannot be figured out without the data. This formalisation of privacy in statistical disclosures has been proven impossible by Dwork et al.,\(^{1458}\) who pointed out how “(t)he obstacle is in auxiliary information, that is, information available to the adversary other than from access to the statistical database”.\(^{1459}\) Open data is exactly that – auxiliary information that is likely to be available to an attacker due to its unrestrained...

---

\(^{1450}\) In this sense, the concern about the protective capacity of anonymization expressed by Ohm seems to be partly overblown (Ohm [n 42]). The point has been made more nuanced in Wu (n 1296).

\(^{1451}\) Ohm and Peppet (n 1254). The extent to which information can be made to relate to a natural person has also been explored by Purtova, ‘The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law’ (n 903); van Loenen, Kulk and Ploeger (n 977).

\(^{1452}\) I refer to the explanation given by Dwork (n 1375). An early definition of semantic security in cryptosystems posits that “(w)hatever is efficiently computable about the cleartext given the cyphertext, is also efficiently computable without the cyphertext” - Shafi Goldwasser and Silvio Micali, ‘Probabilistic Encryption’ (1984) 28 Journal of computer and system sciences 270, 270. In other words, a cryptosystem is defined as semantically secure if an attacker cannot understand anything about the cleartext that he could not understand without the cypher text. Achieving this definition of security – security as confidentiality – can be formally proven. Achieving the equivalent of semantic security in statistical databases, on the contrary, has been formally proven impossible: see Dwork (n 1375). Differently from cryptosystems, where there is (or should be) no auxiliary information other than the cyphertext, when de-anonymizing a statistical database an attacker could have an unknown amount of auxiliary information at his or her disposal that, linked with the statistical database, would allow to infer additional information about particular individuals, such as their identity or a sensitive attribute.

\(^{1453}\) Dwork (n 1375).


\(^{1455}\) In stark contrast with legal notions of privacy, which are characterized for their inherent indeterminacy, so much that e.g. the European Court of Human Rights has routinely declared that it is not possible nor desirable to define what constitutes the right to privacy (rectius, the right to the respect of private life): see inter alia ECtHR, Pretty v. United Kingdom (n 300) §61; ECtHR, Niemietz v. Germany (n 300) §29; Peck v. United Kingdom (n 300) §57.


\(^{1457}\) Dalenius (n 1456) 433.

\(^{1458}\) Dwork (n 1375).

\(^{1459}\) Dwork (n 1375) 2.
accessibility and re-usability, so that its processing is frictionless and presents extremely low transaction costs.\footnote{In a way, open data sharing can have the side effect of lowering the degree of individuals' opacity, one of the main attributes of privacy as framed by Hartzog (n 231); Hartzog and Stutzman, 'Obscurity by Design' (n 287); Hartzog and Stutzman, 'The Case for Online Obscurity' (n 287); Hartzog and Selinger (n 287).}

In this sense, open data is bound to impact individuals' right to personal data protection by increasing the pool of auxiliary data readily available for an attacker, making it theoretically easier to re-identify anonymous data subjects. The material scope of EU data protection law is triggered by the qualification of the information processed as personal data — “any information relating to an identified or identifiable natural person”,\footnote{GDPR, Art. 4.} an ‘identifiable’ person being someone who can be singled out, directly or indirectly, through all the means reasonably likely to be used by the controller or by any other person.\footnote{GDPR, Recital 26.} The extent to which the disclosure of specific kinds of information as open data would be useful to an attacker performing a re-identification attack depends of course from the kind and granularity of the data disclosed; in general terms, there is a necessary trade-off between the utility of an anonymized dataset and the degree of privacy (rectius, confidentiality) it provides.

## 6.5 Beyond the Law: Aligning Policy and Practice

The previous chapter contended that a balance between open data and data protection cannot be achieved due to the very nature of the concepts at stake, so much so that the very idea of balancing open data and data protection is a misnomer. Open data rejects proportionality balancing, while the secondary legislation substantiating the fundamental right to personal data protection is how the rights and freedoms underlying open data are balanced with the countervailing rights and freedoms that may be dented by personal data processing. In a nutshell, personal data cannot be open data unless processed contra legem.\footnote{As discussed in chapter 5, section 5.1, being able to call information ‘open data’ without diluting the term’s meaning requires foregoing the sorts of controls over the information’s accessibility and usability in which fundamental data protection principles, such as e.g. purpose specification, security, and storage limitation, and most of the other norms in which they substantiate. Compliance with data protection, in other words, requires controlling the information’s accessibility, availability, and/or usability, and data that cannot be used by everyone for any purpose is not open by definition. Publishing and re-using personal data as actual open data, therefore, means disregarding many of the core tenets of EU data protection law.} That is not an antinomy, nor an insurmountable issue, as personal data can still be shared and re-used through a less permissive option within the ‘openness spectrum’, i.e. by implementing appropriate access controls and use restrictions. Yet, the only way to reconcile open data per se and data protection without watering down the former’s requirements or breaching the latter would seem to be removing the information in question from the material scope of data protection law by rendering it anonymous before publishing it and allowing its re-use as open data.

This chapter has, so far, tackled the potential over-inflation of the concept of personal data from a legal perspective, addressing it through the tools of doctrinal interpretation. Anonymity is however as much a technical issue and a policy one as it is a normative matter; determining whether a dataset is safe for release as open data involves making sure that data subjects cannot be singled out, that the records it contains cannot be linked to the data subjects they
relate to, and that the dataset does not allow the inference of attributes about a data subject. From this perspective, I believe it would be auspicious to work towards the alignment of data protection law and practice and two other tangential disciplines: information security, on one hand, and statistics (more specifically, statistical disclosure control/limitation), on the other hand.

6.5.1 Aligning Information Security and Data Protection

The ongoing debate about the suitability of anonymity and de-identification as protection mechanisms has been, on the legal and policy side of the issue, quite stale. As Rubinstein and Hartzog noted, “Fifteen years in, the debate has led to polarization, and policy discussions are now splintered. While policymakers [...] have taken note of deidentification’s limits, they have largely ignored developments in adjacent fields [...] Meanwhile, privacy law remains largely unchanged.”\textsuperscript{1464} The core of the issue is that the notion of personal data, on which the entire EU data protection framework stands, is based on a distinction between identifiability and anonymity\textsuperscript{1465} that is as neat in theory as it is unworkable in practice. The status of (big) data as anonymous or personal is often dynamic and contextual, while the law and the relevant jurisprudence treat it as if it were somewhat static.

The anonymity/identifiability dichotomy may be a necessary legal fiction, at least currently, but that should not mean that publishers and re-users of (open) data should live in fear that the assumingly anonymous information they routinely process could turn out to be considered personal. The exegetic tools this chapter has highlighted may reduce the idiosyncrasies deriving from this mismatch between law and practice. Yet, as the amount of available data about individual grows and processing technologies and know-how develop further, it may become necessary for supervisory authorities, and maybe even for the legislature, to provide controllers dealing with data that might refer to identifiable natural persons with more legal certainty about the status of the information they process.

From this perspective, as it has been already proposed,\textsuperscript{1466} I believe it would be appropriate to work towards the (further) alignment of information security and personal data protection. Security, privacy, and data protection have always been inextricably linked.\textsuperscript{1467} Bambauer notes how “(p)rivacy\textsuperscript{1468} establishes a normative framework for deciding who should legitimately have the capability to access and alter information. Security implements those choices”;\textsuperscript{1469} Lysnkey even postulates that the evanescent essence of the right to personal data protection might be sought in data security.\textsuperscript{1470} Data protection legislation has always contained provisions on the security of the processing.\textsuperscript{1471}

\footnotesize
\begin{itemize}
  \item \textsuperscript{1464} Rubinstein and Hartzog (n 1296) 709.
  \item \textsuperscript{1465} See e.g. Article 29 Data Protection Working Party, ‘Opinion 05/2014 on Anonymisation Techniques WP216’ (n 521); Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40).
  \item \textsuperscript{1466} Rubinstein and Hartzog (n 1296) 731.
  \item \textsuperscript{1467} Which is not to say that data security and data protection are identical: security can have different objectives that can be agnostic or even opposed to privacy and data protection: see Lauren Henry Scholz, ‘Information Privacy and Data Security’ [2015] Cardozo Law Review de Novo.
  \item \textsuperscript{1468} To be read as “data protection” in the EU.
  \item \textsuperscript{1469} Derek E Bambauer, ‘Privacy versus Security’ (2013) 103 Journal of Criminal Law and Criminology 667, 669.
  \item \textsuperscript{1470} Lysnkey, \textit{The Foundations of EU Data Protection Law} (n 19) 271.
  \item \textsuperscript{1471} See Convention 108, Art. 7; DPD, Art. 17; GDPR, Art. 32.
\end{itemize}
Anonymity is, to a large extent, comparable to (and often considered as part of) information security, and yet while there is a common understanding of the fact that there is, and cannot be, perfect security, data protection law and policy still rely on the idea that perfect anonymity is achievable. In other words, data protection law, and the identifiability/anonymity dichotomy on which it is based, currently focus on the outputs of anonymisation, i.e. anonymous data, rather than on the process of anonymisation itself. Information security, on the other hand, moves from the assumption that security is a dynamic, fleeting status, and that – since there cannot be perfect security – all one can do is follow standards and best practices, focus on having suitable procedures in place, and consider security as a process, rather than a product. “Data security law is process-based, contextual, and tolerant of harm [...] Because there is no such thing as perfect data protection, data security policy is focused on regular risk assessment, the implementation of technical, physical, and procedural safeguards, and the appropriate response once a system or data set has been compromised”.1473

To be fair, the evolution of data protection in the EU has led to the introduction of several provisions that are oriented towards the development of proper procedures, rather than focusing on the outputs alone. The obligation to perform data protection impact assessments (DPIAs),1474 for instance, and the explicit introduction of norms on privacy by design and by default,1475 or again the provisions about codes of conduct and certifications,1476 exemplify the kind of process-oriented norm I am suggesting. Yet, the determination of what is personal data and what is instead anonymous, which may be difficult when dealing with anonymised microdata, or with big data processing, or again with information relating to a person who is identifiable by an actor other than the data controller, is still eminently output-oriented.1477 “A more sustainable approach would focus on the preconditions and processes necessary for protection”.1478 In other words, it would be desirable for EU data protection law to shift from constructing anonymisation as guaranteed (where “there must in effect be zero risk of an individual being identified within a dataset given whatever assumptions one wishes to underpin the guarantee”)1479 to seeing it as functional1480 and somewhat relative. Suggesting the overhaul of the concept of personal data, and hence of data protection law, is well beyond the scope of this thesis – and, quite frankly, I would not know where to begin.1481 There is a sizeable distance between reckoning that the material scope of EU data protection law should be more process-oriented, contextual, and dependant on the processing’s risk, and implementing those ideas in secondary legislation through a set of rules and principles. Furthermore, I believe there is more value in suggesting a proper course of action within the boundaries of the current normative framework, rather than hypothesising, de iure condendo,

1472 “Irreversible”, in the words of the Article 29 Data Protection Working Party, ‘Opinion 05/2014 on Anonymisation Techniques WP216’ (n 521).
1473 Rubinstein and Hartzog (n 1296) 706.
1474 GDPR, Art. 35.
1475 GDPR, Art. 25.
1476 GDPR, Art. 40 ss.
1478 Rubinstein and Hartzog (n 1296) 729.
1480 See Elliot and others (n 1479) 38.
1481 I refer, on the matter, on the (currently ongoing) work of Dr. Nadya Purtova and of her team in the context of the ‘InfoLeg’ project (http://infolegproject.net/).
what should trigger the material scope of future data protection legislation. And indeed, there is a way to further align information security and personal data protection, with particular reference to the latter’s material scope, and within the boundaries of current EU data protection legislation, i.e. in practice, rather than by amending the law: Statistical Disclosure Control/Limitation.

6.5.2 Data protection and Statistical Disclosure Control/Limitation

The actual issue between open data and data protection, as Chapters 5 and 6 unpacked, is not about balancing, but rather about determining and controlling the latter’s material scope. What is personal cannot be truly open, and what is open cannot be personal, unless the meaning of open data is watered down, or compliance with a high number of data protection rules and principles is neglected. The matter then becomes reconciling the binary conception of anonymity and identifiability that informs the definition of personal data with the fact that, on the ground, the line between personal and anonymous is not as neat as it is on the books. There is, however, a sub-field of statistics that deals exactly with this kind of issue: Statistical Disclosure Control\(^{1482}\) (SDC hereinafter). Statistics, just like any other discipline centred around data processing, has been changing over the past decades: the development of computing technologies has led to the publication of more and more detailed tabular information, and the decentralisation and proliferation of public registers to an ever-increasing availability of microdata.\(^{1483}\) While that had, undeniably, significative positive outcomes for society at large, the ‘data deluge’ and the development of data processing technologies comes with an increased risk of disclosure\(^{1484}\) of individual\(^{1485}\) attributes or identities.

SDC, sometimes also referred to as Statistical Disclosure Limitation (SDL), is the branch of statistics that, simply put, deals with reducing the risk of inadvertently disclosing something that the publication of certain data was not meant to disclose. SDC processes can be used to prevent the inference of the identity of a natural person or of certain attributes concerning him or her, thus rendering anonymous data that could or would otherwise be personal. Indeed, while part of the academic debate has been concerned with the implications of the advances in re-identification sciences and of a number of successful re-identification attacks,\(^{1486}\) other authors have been objecting that SDC as a whole has been progressing as well, and that, while some SDC techniques and methods have not withstood the test of time, many others are still effective.\(^{1487}\) It is also important to note that SDC consists in a broader range of solutions than

---

\(^{1482}\) The term, reportedly, first appeared in Dalenius (n 1456).

\(^{1483}\) I.e. datasets containing several records (e.g. all the family doctors operating within a given area), each of which comprehends a number of variables (or attributes), e.g. age, income, pertaining to the individual statistical units (i.e. each family doctor).

\(^{1484}\) A disclosure is the event where a person or an organisation either recognises or learns something (i.e. the identity or an attribute) about another person, which they did not know already, through the publication of statistical data.

\(^{1485}\) Not just natural persons – a “statistical unit” can also be a legal person, e.g. a company.

\(^{1486}\) See generally Ohm (n 42); Narayanan and Shmatikov (n 1298). Notable examples of successful attacks can be found e.g. in de Montjoye and others (n 621); de Montjoye, Radaelli and Singh (n 621); Rocher, Hendrickx and de Montjoye (n 1298); Culnane, Rubinstein and Teague (n 1298). A systematic literature review has been done by Jane Henrikson-Bulmer and Sheridan Jeary, ‘Re-Identification Attacks—A Systematic Literature Review’ (2016) 36 International Journal of Information Management 1184.

\(^{1487}\) See e.g. Daniel Barth-Jones and others, ‘Assessing Data Intrusion Threats’ (2015) 348 Science 194; El Emam and Alvarez (n 1109); David Sánchez, Sergio Martínez and Josep Domingo-Ferrer, ‘Comment on “Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata”’ (2016) 351 Science 1274 LP; Ann Cavoukian
de-identification and anonymisation alone: not all SDC techniques and methods reduce the risk of identification, inference, and linkage by degrading data quality. Some SDC techniques and methods consist of technical, organisational, and contractual measures that limit the possibility to access, use, or distribute the information they are applied to.

SDC processes are used to determine and prevent the risk of disclosures, and answer to the same issues that data controllers might have, in fringe situations, when determining whether a certain dataset they plan to release still relates to identifiable natural persons, and must thus be processed in compliance with EU data protection law, or can be considered as anonymous, and hence safe for publication as open data. SDC techniques consist of “the set of methods to reduce the risk of disclosing information on individuals, businesses or other organisations”. SDC methods deal with the trade-off between confidentiality and utility by minimising the risk of disclosure to a non-zero, and yet acceptable level. Several different SDC techniques and methods have been developed to deal with different types of outputs.

As discussed in the previous chapter, when dealing with personal data, the choices available to data controllers that wishes to publish such data as openly as (legally) possible are either anonymising it ex ante or choosing a less permissive option within the openness spectrum, so that data protection norms can still be complied with – personal and fully open are mutually exclusive. In cases where the data controller chooses to anonymise a dataset and then release the output of the anonymisation process as open data, rather than foregoing the open data label by implementing the access controls and use restrictions necessary to comply with data protection law, following a SDC process provides a way to do so as effectively.

The purpose of SDC is to impede that confidential information that might be inferred from a data set – i.e. an unknown attribute about a known subject or the identity of an unknown subject – is connected to a specific person, or statistical unit. SDC is thus concerned with preventing the risk of disclosures for statistical units, which may be both natural and legal persons, rather than for data subjects, and is meant to impede the disclosure of a broader range of information than personal data, e.g. business data, or sensitive information about...
groups. The difference between the notions of statistical unit and data subject, on one hand, and of confidential data and personal data, on the other hand, does not however seem to matter: data subjects are a subset of the entities that can fall under the definition of statistical unit, and personal data are a subset of what can be deemed as confidential for SDC purposes. In other words, even if the purpose and scope of SDC is different than personal data protection’s, using SDC methods to determine and control whether a certain data set falls under the material scope of EU data protection law is well within SDC’s purposes.\textsuperscript{1494}

SDC considers information that may or may not need to be sanitised for different reasons, including personal data protection, and – through the analysis of the dataset, the definition of the disclosure risks, and the selection of appropriate SDC techniques and methods to tackle them – ends with an output that is safe according to the parameters set. Those parameters (the definition of disclosure, of disclosure risks and scenarios, and the likes) can be tuned to meet the normative threshold set by EU data protection law. The application of a SDC process to the (fringe) situations where a controller is in doubt about (or wants to control) whether a dataset it plans to publish (still) contains information relating to identifiable individuals, provided suitable parameters and SDC methods are chosen, appears to be the best controllers can do to overcome the indeterminacy of the notion of personal data.

6.6 Implications for the Pursuit of a “Balance” between Open Data and Data Protection

The research question from which this thesis originates postulates that there is something wrong with the material scope of EU data protection law, which would be indeterminate and broad enough to encompass kinds of information traditionally deemed non-personal or that have been anonymised by the controller. That, in turn, would compress the possibility to share and re-use information as open data, which would purportedly have dire consequences on the development of the ‘smart cities’ of the future, whatever those may be. As readers might have gathered, I believe none of that is true as such. The considerations about the material scope of EU data protection law above\textsuperscript{1495} may mitigate the indeterminacy and breadth of the notion of personal data. The concerns about the fact that all information may relate to somebody,\textsuperscript{1496} and that everyone may be identifiable,\textsuperscript{1497} seem to have been blown out of proportion when seen in light of EU data protection law. At the same time, data protection legislation is built with a certain degree of flexibility and risk tolerance, so that the detrimental effects of an extensive or overly zealous application of its individual provisions can be weathered by the system as a whole. All in all, while the identifiability and the ‘relating to’ requirements, on their own and in a vacuum, may be overly expansive, the notion of personal data as a whole is not.

That holds particularly true if coupled with the lessons from information security discussed in the previous subsection. While it is commonly accepted that there cannot be perfect security, EU data protection law is however (still) modelled on the assumption that there can be perfect

\textsuperscript{1494} A mapping of some SDC sanitation techniques and contextual controls to the categories of data in the GDPR has been performed e.g. in Hu and others (n 1332).

\textsuperscript{1495} See s. 6.3.

\textsuperscript{1496} See Ohm and Peppet (n 1254); Purtova, ‘The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law’ (n 903).

\textsuperscript{1497} See Ohm (n 42).
anonymity. That may be a necessary legal fiction, as the material scope of EU data protection law is based on the distinction between anonymous and personal. It may however also be just a burdensome legacy that derives from the fact that EU data protection legislation is still modelled after laws drafted between the ‘70s and the ‘90s, when the state of the art did not allow for the same ease of identification as it does today, and when the amount of digital information relating to individuals was a fraction of what is now available. It would arguably be a step forward, as it has been suggested elsewhere, to follow information security – by law, policy, or by means of interpretation – in its contextual, risk-tolerant, and process-oriented view of data protection.

Be it as it may, there can still be fringe situations where data controllers are unsure about whether a de-identified dataset is truly anonymous, and thus safe for release as open data, or whether natural persons can still be identified within or through the records it contains. The issue is however not new, but has rather been tackled extensively by the field of statistics. Implementing a proper SDC process, rather than merely removing direct and indirect identifiers, is supposed to take care of the fringe situations where a personal data controller that decides to release anonymised data as open data (rather than publishing personal data through a less restrictive data sharing policy) may be in doubt about the qualification, as anonymous or personal, of the information whose publication is planned. Data controllers cannot do much more than that, they should not be required to, and – within the current legal framework – they are arguably not.

The real issue is not about personal data overinflating and encompassing more and more information traditionally seen as anonymous, which would suggest the need of curtailing the material scope of EU data protection law. It is rather about the fact that there is an unavoidable trade-off between privacy and utility, and between openness and protection. If the publication plan is to release as open data a certain dataset that derives from another dataset containing personal data, the former’s utility should be degraded so that to whom the original data refer can no longer be singled out, have their attributes inferred, or have their records linked. If the information’s utility is a paramount concern, and cannot be lowered to the point where anonymity is (reasonably likely to be) guaranteed, then open data is not a suitable data sharing policy. That does not necessarily mean that such (personal) data cannot be openly accessible or machine readable, but it does mean that it cannot be usable and re-usable by anyone and for any purpose, lest data protection core tenets are infringed.

1498 See generally Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU (n 19); Mayer-Schönberger (n 679); Hondius, Emerging Data Protection in Europe (n 694).

1499 Rubinstein and Hartzog (n 1296).

1500 See generally Hundepool and others (n 1491); Leon Willenborg and Ton De Waal, Elements of Statistical Disclosure Control, vol 155 (Springer Science & Business Media 2012); George T Duncan, Mark Elliot and Juan-José Salazar-González, Statistical Confidentiality: Principles and Practice (Springer New York 2011). See also Elliot and others (n 1479).

1501 See Wu (n 1296).

1502 “(O)pen data environments are really only appropriate to data that are either a pesso nal or have been through an extremely robust anonymisation process that ensures with a very high degree of confidence that no individual could be re-identified and no statistical disclosure could happen”: Elliot and others (n 1479) 55–56.

1503 E.g. the purpose limitation principle.
6.7 CONCLUSION

The combination between the low identifiability threshold and the multiple ways in which information can relate to natural persons render the notion of personal data potentially very broad. Individuals are ubiquitously tracked through the devices they carry, and advances in technology and research make identification a progressively easier endeavour. At the same time, information can be linked to (identifiable) natural persons in many ways – i.e. through its content, its purpose, or its result. The notion of personal data, and the identifiability/anonymity dichotomy on which it is based, can thus be seen as overly fuzzy and expansive concepts. Such concern is more than legitimate from a regulatory perspective, thinking de iure condendo. However, focusing on it while determining, de iure condito, what constitutes processing of personal data, can lead to perverse results. Law, jurisprudence, and doctrine, nonetheless, provide enough exegetic tools to ensure that the concept of personal data remains flexible enough not to deprive individuals from their right to data protection, while still sufficiently narrow as not to cover all data all the time.

I have argued that the potential over-inflation of the concept of personal data within the current regulatory framework is due, in particular, to the possibility to interpret too extensively the identifiability requirement and the meaning of ‘relating to’. I have also argued, however, that the law and the related doctrine and jurisprudence also contains elements that can permit to avoid an overly extensive interpretation. The GDPR embeds a certain degree of flexibility in its application, which reflects also on the notion of personal data and on the elements that compose it. The potential width of the ‘relating to’ link is tempered by its interaction with the identifiability requirement, particularly if one considers (personal) data within its lifecycle, and the fact that EU data protection law is meant to regulate the processing of individual attributes too, not just identifiers. On the other hand, the expansive effect of the identifiability requirement can be reined in by applying the plethora of criteria devised to perform the ‘reasonable likely’ test not only to the means that can be used to identify the data subject, but also on the persons to whom those means can be available, and on their relationship with the controller. Despite its indeterminate and relative character, I believe that the notion of personal data can still be a viable basis for policy and legislation. A nuanced epistemological approach to determining what constitutes personal data, through the appropriate interpretative criteria, either mitigates the dreaded effects of an overextension of the material scope of EU data protection law, or justifies the rationale for which information should be considered personal.

The adoption of an evolutionary interpretation is however a precondition for the concept of personal data to keep being a suitable regulatory mechanism in an ever-changing technological landscape. To this end, it would be beneficial for personal data protection to learn from information security the fact that, as there cannot be perfect security, there cannot be (any formal guarantee of) perfect anonymity either. Data protection law, and the determination of its material scope, should become more contextual, risk-tolerant, and process-oriented. With particular respect to the latter desideratum – process-based data protection – it is also important to highlight how, while personal data protection law has just begun coming to terms with the non-absolute nature of anonymity and anonymisation, statistics have been dealing with this very issue since before SDC began to be referred to as such. The SDC process outlined above, and the application of suitable SDC methods amongst the many available in literature, should provide data controllers with enough tools to deal with the fringe situations where they
are unsure about whether the data they (plan to) process (still) refers to identifiable natural persons.

What does this all mean for open data? For starters, it mitigates the concerns about traditionally non-personal (i.e. anonymous) data suddenly becoming personal by virtue of an unfortunate combination of normative overreach, technological developments, and black magic. While all data may be relatable to a natural person or another, it being actually related to an identified or identifiable individual is a different thing. Entities processing weather data\(^{1504}\) or maps\(^{1505}\) may be reassured in knowing data protection law and data protection lawyers will not be coming for them anytime soon. The widespread concern about the relative nature of anonymity\(^{1506}\) and about the vulnerability of anonymisation\(^{1507}\) is surely well grounded, but so are the statistical disclosure control methods and techniques that have been developed, and that are still being developed, as a response.

EU data protection law might be lagging behind in adapting (the interpretation of) its material scope to the state of the art in technology and know-how, but that is somewhat marginal. The core of the issue is that there is an inherent trade-off between identifiability (and hence personal data protection) on the one hand, and ‘openness’ and granularity (and hence utility), on the other hand.\(^{1508}\) “Balancing” open data and data protection – rectius, balancing data protection and the rights and freedoms underlying open data – entails acknowledging how publishing a dataset deriving from personal data inevitably requires either degrading its granularity until anonymous, or restricting its accessibility and/or usability in compliance with secondary data protection legislation.

---


\(^{1505}\) Cfr. van Loenen, Kulk and Ploeger (n 977). Thematic maps, who overlay topographic maps with (the result of the processing of) other data, can still be personal data, depending on the level of granularity and the kind of information represented over the topographic layer. Location data under the definition of the ePrivacy Directive is of course a whole other thing.

\(^{1506}\) Rubinstein and Hartzog (n 1296).

\(^{1507}\) Ohm (n 42).

\(^{1508}\) See Wu (n 1296).
7. CONCLUSIONS

The postulates from which the research proposal motivating this thesis moved maintained that smart cities need open data, and open data sharing and re-used is getting increasingly curtailed by personal data protection norms, and that the legislature ought to do something to balance open data and data protection, lest we forego the oft-touted benefits of the cities of the future. It makes sense, intuitively: the instrumentation and datafication of the built environment is a matter of fact, and the trend towards the conflation between code and space is on the rise. The plasticity of the notion of smart city makes for an easy alignment with open data. On one hand, open data can be seen as an output of the myriad of reconfigurable sensors that increasingly populate the urban environment. On the other hand, open data can also be seen as a component of the information revolution that has been sweeping through urban sciences, and that is often considered as a part of – or equated with – the idea of ‘smart cities’.

At the same time, as digitalisation, computing, and networking grow in weight in modern society, the rights to privacy and to personal data protection are mounting in importance and widening in scope. The core tenets of data protection secondary legislation (e.g. purpose specification and limitation, data minimisation, security and storage limitation) are fundamentally incompatible with the requirements that must be respected to qualify information as (actual) open data – fully machine-readable information that is freely accessible, usable, distributable, and re usable by anyone and for any purpose. As a consequence, it would seem legitimate to be concerned that, as the material scope of data protection expands, the amount of information accessible and reusable as open data will proportionally decrease.

As you might have gathered, I do not believe that to be entirely true – at least not for any value of ‘true’ that would require the modification of the current data protection regime through regulatory means. My research convinced me that the smart city is a socio-technical

---

1509 See e.g. Caragliu, Del Bo and Nijkamp (n 17); Nam and Pardo (n 113); Nam and Pardo (n 81); Townsend (n 6); Goerge (n 206); Steven E Koonin and Michael J Holland, ‘The Value of Big Data for Urban Science’ in Julia Lane and others (eds), Privacy, big data, and the public good: Frameworks for engagement (Cambridge University Press 2014).
1510 Kitchin and Dodge (n 10).
1511 See e.g. Open North (n 33); Hielkema and Hongisto (n 648); Walravens, Breuer and Ballon (n 14); Bartenberger and Grubmüller (n 648).
1513 Kitchin, The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences (n 4); Schwab (n 817); Floridi (n 817).
1515 See e.g. van Loenen, Kulk and Ploeger (n 977); Purtova, ‘The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law’ (n 903); Brkan, ‘The Unstoppable Expansion of the EU Fundamental Right to Data Protection: Little Shop of Horrors?’ (n 903).
1516 Open Knowledge International (n 32).
1517 See Pollock (n 635).
construct, malleable enough to be able to encompass whichever narrative manages to include both the *topoi* of urbanities and of technology. It also highlighted how open data is a means, rather than an end, fit to foster a plethora of rights and freedoms, but nor a right on its own. The theory about the substance of the right to personal data protection adopted in this thesis also underlines how it is a tool of transparency, meant to allow processing while channelling it through a set of rules, rather than to curtail it.

The issue is that, despite the permissive nature of data protection, many of the rules and principles in which it substantiates represent forms of access control and use restriction. Those, by definition, rule out the possibility to define a (personal) data set as open data, at least not without watering down the actual meaning of open data so that it encompasses less permissive sharing and re-use policies. Yet, while its social and economic value is apparent, open data is not the only way of publishing and re-using information to harness its secondary value. Indeed, whenever the rights and freedoms that can be supported through open data sharing and re-use clash with countervailing rights and freedoms, the outcome of a (proportionality) balancing test cannot be open data by definition. Open data is the most extreme data sharing and re-use option, and its requirements reject the mechanisms through which proportionality in (personal) data sharing and re-use is pursued, and balancing performed.

In other words, data protection and open data are mutually exclusive, intended to run on parallel tracks, and not to intersect. In cases where there is the chance that they do (i.e. when the data considered may relate to indirectly identifiable natural persons) the core of the matter then becomes delineating a distinction between what is personal and what is not. The issue is not really ‘balancing’ open data and data protection – that, as chapter 5 explained, cannot really be done. The issue is for data controllers to be able to determine, with sufficient confidence, what is personal and what is anonymous, so they can implement the appropriate technical and organisational measures for the former, and confidently publish and re-use the latter as open data.

Concerns about the overextension of the concept of personal data, and thus of the material scope of EU data protection law, are certainly not without merit. Yet, the issue of the interaction between personal data and open data is not one of scope creep: while the concept of personal data is (meant to be) broad, the problem has rather been the opposite, the fact that the applicability of EU data protection law has often been resisted by data controllers. The issue is one of boundaries: there may be fringe situations where data controllers of a varying degree of sophistication are either unsure about whether the data they process is (still) personal or anonymous, or mistaken about it. Those are, however, a very small subset of all (personal) data processing instances, and there are routes that can be taken – in the law and in its interpretation, in policy, and in practice – to reduce the degree of legal and practical uncertainty that may arise.

---

1518 See Sadowski and Pasquale (n 140); Sadowski and Bendor (n 226); Söderström, Paasche and Klauser (n 173).
1519 De Hert and Gutwirth (n 858); De Hert and Gutwirth (n 236).
1520 Both in the sense of ‘apersonal’ and in the sense of anonymised.
1521 See e.g. Purtova, ’The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law’ (n 903); van Loenen, Kulk and Ploeger (n 977); Ohm (n 42); Ohm and Peppet (n 1254); Brkan, ’The Unstoppable Expansion of the EU Fundamental Right to Data Protection: Little Shop of Horrors?’ (n 903).
1522 See e.g. Google Spain (n 1118).
This chapter concludes the thesis. It summarises the previous chapters’ main findings, acknowledges their limitations, and suggests further avenues for research.

7.1 SMART CITIES AND RED HERRINGS

The starting point of this research was the notion of smart city, and the fact that the apparent clash between open data and personal data protection might hamper its development. The original idea was thus to figure out what a smart city is, what does a city need to be defined as such, and how would data protection be in the way of the fulfilment of its requirements. Yet, the syntagm “smart city” does not indicate a specific technology or set of technologies – although some are staples in the smart city imaginary. The smart city, rather, can be seen as a technological paradigm: “the epistemological definition as an ‘outlook’, a set of procedures, a definition of the ‘relevant’ problems and of the specific knowledge related to their solution”. As Dosi noted already several decades ago, “each “technological paradigm” defines its own concept of ”progress” based on its specific technological and economic trade-offs”. In other words, the requirements for ‘smartness’ are not inherent to the concept of smart city, but are rather set by the political agenda of the relevant stakeholders. There are no ‘fixed’ smart city requirements, just the ones set by the polity of reference.

The term ‘smart city’ can also refer to the development in urban sciences, to the kind of ‘smart’, data-driven, networked urbanism that has been made possible by technological development, by the rise of computing, and by the data deluge that followed. This thesis does not underestimate the effect that the data revolution has had on urban scholarship, management, and governance. Yet, the ‘smart city’ can also be framed as a mere socio-technical construct: a (largely corporate-driven) narrative based on the implementation of technological solution(ism)s to the extremely multifaceted challenges faced by modern urbanities. From this point of view, the term ‘smart city’ seems to answer to needs of marketing and obfuscation: ‘smart city’ may be just the portrayal of a better future through the deployment of a given set of technologies, meant to crowd out alternative visions of urban development while pushing forth the agenda that underlies the narrative. The diversity in the framing of the different smart city initiatives that are popping up all over the globe highlights how there is not a unitary answer to what being ‘smart’ means for cities, bar the deployment of new technologies in the urban environment.

Open data fits seamlessly with the smart city narrative: it can be seen as both a by-product of the sensors with which the built environment is instrumented, and as an enabler of the “smart

---

1523 Broadly speaking: sensors and actuators, cloud computing infrastructures, and big data analytics: see Edwards (n 7); Finch and Tene (n 21). See also Future of Privacy Forum (n 321).
1525 Dosi (n 1524) 148.
1526 Kitchin, ‘Data-Driven, Networked Urbanism’ (n 3).
1527 See Sadowski and Bendor (n 226).
1528 See Söderström, Paasche and Klauser (n 173); Hollands (n 175). See also Sotirios Paroutis, Mark Bennett and Loizos Heracleous, ‘A Strategic View on Smart City Technology: The Case of IBM Smarter Cities during a Recession’ (2014) 89 Technological Forecasting and Social Change 262; Alan Wiig, ‘IBM’s Smart City as Techno-Utopian Policy Mobility’ (2015) 19 City 258; McNeill (n 174).
1529 Morozov (n 230).
1530 See e.g. Sadowski and Bendor (n 226); Söderström, Paasche and Klauser (n 173).
revolution”. Yet, defining it as a requirement for the development of the smart city seems to answer to a particular policy agenda, rather than to open data being an actual condition sine qua non for future urbanisation. While certainly beneficial to urban development, be it smart or not, open data is just one amongst the many data sharing and re-use options available. Data protection, on the other hand, has the status of fundamental right in the EU, and is guaranteed by secondary legislation: its relevance to the smart city discourse and to the open data ecosystem is not up for debate. Quite the contrary: personal data protection is inherently threatened by the instrumentation and datafication of the built environment, and arose as a consequence of the emergence of the computing and networking technologies on which the smart city is based.

Be it as it may, with respect to this thesis’ main area of inquiry – the idea of balancing open data and data protection – the reference to the smart city is a red herring. Data protection is intrinsically at odds with open data, regardless of whether the context of their interaction is the smart city or not. One might argue for the specificity of the smart city by contending that the instrumentation of the built environment might allow for new kinds of data to be collected, which could then be shared and re-used as open data, if it were not for those pesky data protection norms. I reckon that person would however be hard-pressed to point out specific categories of information whose collection has been enabled by the ‘smart city’ (a), that are personal data (b), and that it would not be entirely wanton to release as open data without rendering them anonymous beforehand (c). Ultimately, the “new” kinds of information collected in the smart city environment are either anonymous from the outset (e.g. environmental data) or involve (the tracking of) individuals, be it through the processing of their biometric data (e.g. face or gait recognition) or the monitoring of their devices (e.g. Wi-Fi monitoring, or smart metering). In the former case, its publication and re-use as open data is not hampered by data protection legislation at all; in the latter, the idea of sharing it as open data without performing a proper anonymisation procedure is so disproportionate that it results abhorrent: the idea of open data should be about open cities, rather than open citizens.

We may accept to see the smart city as a technological paradigm, rather than a narrative or a socio-technical construct. In that case, it seems significant to note how, in the face of technological development that creates or increases power and information asymmetries between data controllers and data subjects, the regulatory response of the EU has been tightening the rules, rather than deregulating the matter. Even if we were to frame the idea of balancing open data and data protection within the specific context of the smart city environment, the push toward lowering the degree of protection afforded by EU data...
protection law would surely be resisted by a specular moment. In other words, the option of regulating smart city technologies by virtue of their perils is as legitimate as the idea of deregulating data protection (i.e. “balancing” it with open data) for the pursuit of the smart city’s promises, perhaps even more. This thesis however leans towards seeing the smart city as a socio-technical construct. The plasticity of the smart city narrative permits to frame whichever initiative is even tangentially connected to both technology and the urban environment as a smart city development. A spin on the wording used is well enough to frame, for instance, E-health, public and private transportation services, or utilities supply as part of the smart city environment. From this point of view, I contend that there is little to no added value in considering the idea of balancing open data and data protection within the smart city context specifically. If whatever involves technology and urbanities can be made to fall under the smart city umbrella, it is arguable whether the smart city narrative can provide a reason to shift anything in the balance between open data and data protection.

Finally, data protection law is generally meant to be technology neutral. Hildebrandt and Tielemans distinguish between three objectives of technological neutrality. The first is “compensation”, which refers to the need to promulgate technology-specific regulation when new technologies or technological paradigms threaten human rights. The second is “innovation”: the need to prevent biases for or against specific technologies in ways that stifle innovation and growth. The third is “sustainability”, which refers to the need to enact legislation at a level of abstraction that prevents it from becoming outdated too soon.

It is disputable whether the smart city, as a narrative and a socio-technical construct, engenders the need for compensation, particularly if that entails compressing the right to data protection. For starters, the technology on which the smart city narrative rests are already covered by a comprehensive normative framework. Furthermore, smart city technologies (cloud infrastructures, networked sensors, and big data analytics) are naturally prone to bolster the rights and freedoms underlying open data, on the one hand, and to interfere with individuals’ rights to privacy and data protection, on the other hand: not vice versa. Even if there were the need for regulatory compensation, such compensation would hardly entail watering down personal data protection to further open data. Similarly, the concept of smart city per se is not precise enough to justify a targeted regulatory intervention aiming at ensuring, in pursuit of the “innovation” objective, the law’s neutrality towards specific technologies. Likewise, the competition between open data and data protection does not seem to be prejudicial to innovation in the smart city domain specifically. Finally, the law is meant to be sustainable, generic enough to withstand the test of time, and much of the applicable one has just recently been overhauled: with particular reference to the interaction between open data and data protection, there is nothing in the notion of smart city that indicates the need of additional regulatory specificity.

All in all, it does not seem that resorting to the smart city construct can provide enough reasons, if any, to shift the balance between open data and data protection. The smart city is a narrative: does not have stricto sensu requirements, just policy objectives. The technological layer on which the social construct is based is as favourable to furthering the rights and

---

1539 See e.g. Galdon-Clavell (n 46); van Zoonen (n 139); Finch and Tene (n 21); Edwards (n 7).
1540 See Koops (n 344); Reed (n 344).
1541 Hildebrandt and Tielemans (n 344).
1542 Edwards (n 7).
freedoms underlying open data as it is inherently threatening to individuals’ right to privacy and data protection: there is no reason for a reference to the smart city to change the balance between the two. The smart city narrative, finally, does not challenge the neutrality of the current regulatory framework, at least not with respect to the interaction between open data and data protection.

7.2 OPEN DATA, PSI, AND PERSONAL DATA

Regardless of whether it is framed in the smart city context or not, the idea of balancing open data and data protection implicitly assumes that open data is some sort of right on its own. Open data, however, is not a right: it is an idea and an ideal, a movement, and a data sharing policy and licensing scheme, which may be beneficial to the furthering of a broad array of rights and freedoms, but is not one of them on its own. It is those rights and freedoms (e.g. freedom of information, access to documents, freedom of business) that may need to be weighed against the right to personal data protection. Resorting to the smart city narrative, which is no stranger to the misconstruction of what the existence of a right entails, might obfuscate the fact that open data is a means, and not an end on its own. Yet, upon closer examination, it is clear how the idea of balancing open data and data protection rests on a false equivalence: simply put, open data is not a right, and data protection is.

The right to re-use that has been granted by the PSI Directive since 2013, on the other hand, is indeed a right in the technical meaning of the term. The idea of balancing open data with personal data protection might thus be rather seen as the pursuit of a balance between data protection and the right to PSI re-use. While that takes care of the idiosyncrasy of the idea of balancing a fundamental right (data protection) with a data licensing scheme (open data), the idea of balancing data protection with the right to PSI re-use also presents three issues on its own.

First, data protection is a fundamental right, enshrined in the Charter and in the ECHR, and substantiated by secondary EU law, while PSI re-use is granted by the PSI Directive – EU secondary legislation. Hierarchical issues aside, as one is a fundamental right and the other is not, comparing PSI re-use and data protection requires operating at different levels of abstraction. Personal data protection is ensured by a number of sub-rights – e.g. the right to object, or to be informed about the processing, or the right to erasure – enshrined in secondary legislation: it is those rights that interact with the right to re-use. At a (quasi)constitutional level, data protection would have to be balanced with the rights and freedoms underlying the right to PSI re-use.

Second, open data and PSI are not (necessarily) the same thing, despite the fact that PSI legislation has increasingly been associated with the popular concept of open data since the revision of the 2003 PSI Directive, up to the point where the term ‘open data’ is referenced in

---

1544 See Paolo Cardullo, Cesare Di Felicianonio and Rob Kitchin, The Right to the Smart City (Emerald Publishing Limited 2019). “Smartness” aside, with particular reference to the city, the term “right” has a fairly long history of dilution – see Henri Lefebvre, Le Droit à La Ville (Paris Anthropos 1968).

1545 See Janssen and Hugelier (n 434); Katleen Janssen and Sara Hugelier, ‘Open Data: A New Battle in an Old War Between Access and Privacy?’ in Mireille Hildebrandt, Kieron O’Hara and Michael Waidner (eds), Digital Enlightenment Yearbook 2013: The Value of Personal Data (IOS Press 2013).

1546 Charter, Art. 8.

1547 ECHR, Art. 8.
the 2019 recast of the PSI Directive’s very title. On one hand, while open data is mean to promote innovation and economic growth, public sector efficiency, and governmental transparency and accountability, EU PSI legislation has an eminently market-making underpinning. Equating PSI and open data would discount the facets of open data that are not connected with growth and public-sector efficiency,¹⁵⁴⁸ but rather with the right to access or freedom of information, each of which is covered by a legal regime other than PSI.¹⁵⁴⁹ On the other hand, EU PSI law allows to set conditions on the publishing and re-use of documents that are not necessarily compatible with the open data requirements – at least, not with a strict definition of open data. In other words, PSI licenses can contain licensing terms that are not open data-compatible, as they restrict the data’s accessibility and usability. Those conditions,¹⁵⁵⁰ most of all, while incompatible with open data’s requirements, permit to comply with the tenets of EU data protection law, basically eliminating the (supposed) points of conflict motivating the research questions underlying this thesis.

Third, the interaction between PSI and data protection is, by now, an established part of EU law,¹⁵⁵¹ and is regulated by a clear set of norms which do not, ultimately, impede the processing of the information to which they apply. If we were to equate open data and PSI, there would not be the need to balance competing rights and freedoms: the operation to be performed would be the simple subsumption of a case (the publication of PSI that are or contain personal data) under a set of rules (i.e. the data protection exemption, or the non-prejudice clause of the PSI Directive). The idea of “balancing” data protection and PSI re-use, from this perspective, just equates to renegotiating their established relationship, in detriment of data subjects and to the advantage of PSI re-users, in this case by resorting to the popularity of the smart city imaginary.

Ultimately, the right to PSI re-use is not a right to open data. Regardless of its nominal alignment with open data, PSI maintains an eminently market-making underpinning, and should not be seen as a proxy for rights and freedoms that are upheld through different legislative means in the EU legal system.¹⁵⁵² Conversely, those other rights and freedoms (e.g. access to documents, freedom of information) should not be used as a Trojan horse to advance the predominantly economic interests advanced by PSI law, particularly to the detriment of individuals’ right to data protection.

In that respect, the current legal system already provides for a clear and established set of rules regulating the interaction between PSI re-use and data protection: ad hoc exemptions,¹⁵⁵³

¹⁵⁴⁸ Zuiderveen Borgesius, Van Eechoud and Gray (n 41); Scassa (n 614).
¹⁵⁵⁰ E.g. licensing conditions prohibiting reidentification, or binding the processing to certain purposes; technical and organisational security measures; internal and inter-institutional access policies.
¹⁵⁵³ Art. 1(2)(h) of the 2019 PSI Directive.
and a non-prejudice clause. The same regulatory mechanisms seem to be employed in all the legislative instruments that substantiate the other rights and freedoms underlying open data; personal data are exempted from the applicable regime, and the heightened ‘openness’ that the law mandates is explicitly meant not to be prejudicial to data protection. Indeed, the respect for individuals’ rights to privacy and data protection, and the exclusion of personal data from the scope of open data initiatives, might be more readily seen as an implicit requirement of open data sharing and re-use, rather than an obstacle to its development.

7.3 DATA PROTECTION AND ITS DETACHMENT FROM PRIVACY

The perception of the necessity, or even the opportunity, of pursuing a balance between open data and data protection, seems to derive, to a large extent, from a misconception of the right to personal data protection, and of the mechanics through which it is ensured. The sentiment motivating the research question this thesis is meant to answer was that data protection would be straying away too much from its privacy roots, expanding its scope beyond its original function, and curtailing information flows that are foundational for the development of smart cities. While it is true that data protection evolved away from privacy, and that its prominence in modern society has been growing on par with the one of computing and networking technologies, understanding how and why it happened leads to justify and cherish the emergence of a new fundamental right, rather than regretting its departure from privacy. In turn, understanding how data protection secondary legislation works leads to reframe the dichotomy between personal data and open data: the choice is not really between open and closed data, but rather between open data and other options within the openness spectrum.

The growth in pervasiveness of computing technologies in contemporary society gradually engendered a regulatory response, which happened to take the form of the right to personal data protection. Data protection spun out of privacy, and it is still called ‘data privacy’ or just ‘privacy’ in jurisdictions other than the EU. Yet, at this point in time, data protection is more akin to a framework regulating computing and networking than to one protecting individuals’ right to private life. It is true, in other words, that data protection strayed away from its original privacy framing; that is, however, a good thing, and not by itself a reason to argue that data protection should be brought back to its past privacy dimension. Regulation evolves as society changes, and society changed quite a bit since data protection began to move away from privacy, half a century ago by now. If data protection law has to change—which is itself arguable, given the recent overhaul—it is not by scaling back to being a facet of privacy, but rather by changing incrementally and moving forward.

If data protection were to be seen (by now both anachronistically and contra legem) as a facet of privacy, it may comprehensible how certain stakeholders can perceive the fact that some

---

1554 Art. 1(4) of the 2019 PSI Directive.
1555 Aside from Art. 1(2)(h) and 1(4) of the 2019 PSI Directive, see e.g. Art. 4(2)(f) of the AEI Directive, and the last period of Art. 4(2). See also, for instance, Art. 10 of Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010 on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport, 6.8.2010, OJ L 207/1, which deals with both data protection and PSI re-use in the ITS domain.
1556 Bygrave, Data Privacy Law: An International Perspective (n 269) 23 ss.
1557 See e.g. Koops (n 1301).
1558 De Hert and Papakonstantinou (n 802); De Hert and Papakonstantinou (n 807).
categories of information are personal data – cadastral parcels, for instance – as being symptomatic of the overextension of the material scope of EU data protection law. Yet, data protection has become, at its very core, an attempt at the regulation of computing. From this perspective, it is easier to understand that information about things but that relates to individuals by virtue of its purpose or of the result of the processing, for instance, or information relating to an individual that is indirectly identifiable by a person other than the controller, is personal data, and is meant to be personal data.\textsuperscript{1559} The fact that personal data protection came to encompass information that is not traditionally seen as involving people’s private lives is not a reason to wish it back to its privacy dimension, but rather a sign of its emergence as a separate and autonomous fundamental right.

On account of the history of the materialisation of data protection as an independent fundamental right in the EU, of the characteristics of the secondary legislation that substantiate its fundamental right dimension, and of the relative jurisprudence, it seems that the most modern, rational, and coherent view of data protection is a ‘proceduralist’ one. Data protection has become, and perhaps always was,\textsuperscript{1560} a potpourri of procedural rules, directed mostly at data controllers, that channels the activity of (personal) data processing through a set of best practices which aim at lowering the likelihood and impact of informational harm to data subjects. The (quasi)constitutionalisation of data protection operated by the Charter is not a regulatory quirk, but rather the explicit acknowledgement of a ‘new’ fundamental right, unequivocally distinct from privacy, meant to operate as a shield for individuals against the perils of badly engineered information processing architectures.

From this perspective, bringing back data protection to its privacy past would only serve to muddle and compress a regulatory framework that is not (only) meant to protect seclusion, self-determination, and all the other many values underlying privacy, but rather to tackle the power and information asymmetries that may arise from contemporary computing practices. The justification for doing so, according to the idea from which this thesis moves, would be that data protection (a fundamental right) clashes with open data (a data sharing policy, or a set of licensing conditions) to the detriment of ‘smart cities’ (a socio-technical construct, to be nuanced, or a neoliberal narrative, to be forthcoming). A different framing of the problem statement would substitute ‘open data’ with ‘the right to re-use PSI’, in which case the reason for ‘bringing data protection back to privacy’ could be paraphrased into the fact that data protection clashes with the facet of freedom of business that is fostered by PSI re-use,\textsuperscript{1561} again to the detriment of the fabled smart city. Those are both, to my understanding, statements that belongs to the realm of policy and politics, rather than to the law’s: there are no real normative reason for doing so, bar the legislature’s whim and value judgement.

A second misunderstanding, aside from the conflation between privacy and data protection, relates to the mechanics of data protection secondary legislation. The framing of the research question underlying this thesis presents open data as a binary alternative to keeping information siloed in. However, as opposed to privacy, data protection is a fundamentally permissive regime,\textsuperscript{1562} where the default is that personal data processing is allowed, albeit under a set of rules and principles that curb the power and information asymmetries running

\begin{itemize}
\item \textsuperscript{1559} Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40).
\item \textsuperscript{1560} See the description of the first waves of national data protection legislation in Europe described in Mayer-Schönberger (n 679). See also Hondius, Emerging Data Protection in Europe (n 694).
\item \textsuperscript{1561} See s. 3.3.1 and 3.4.2 above.
\item \textsuperscript{1562} See e.g. Fuster and Gutwirth (n 882); De Hert and Gutwirth (n 858); De Hert and Gutwirth (n 236).
\end{itemize}
between controllers and data subjects. In other words, the framing of the open data/personal data dichotomy for which data protection legislation hampers ‘smart city’ developments portrays an alternative between open and closed data that does not account for the other shades of the openness spectrum.\textsuperscript{1563} If one considers the extent to which EU data protection law permits personal data processing, it results clear that the choices to be made when setting up a processing architecture are mostly about how to process personal data, rather than whether. From this point of view, the idea of ‘balancing’ open data and data protection is not about enabling data sharing and re-use in situations where it would otherwise be precluded, but about watering down, or removing tout court, the (largely procedural) rules and principles governing personal data processing.

7.4 On balancing

The issue becomes clearer when considering what ‘balancing’ signifies from a normative perspective.\textsuperscript{1564} The thought of balancing open data and data protection, be it for the benefit of smart cities or not, intuitively evokes the idea of harmonising competing interests and values, so that the outcome may reflect a better equilibrium between them. Yet, that is a misleading picture, as open data is not an interest or a value on its own, but a data sharing and re-use policy substantiated by a set of licensing conditions. The place of open data in a balancing exercise involving competing rights and freedoms is not as an input – one of the values to be balanced – but rather as a possible output of the balancing test. When reframed this way, the matter becomes assessing whether, when individuals’ right to personal data protection competes with the rights and freedoms that may be advanced through open data sharing and re-use, open data can be a proportionate outcome. That involves assessing whether there can be, within the realm of the legally and factually possible, a compression of the right to data protection stark enough to permit the sharing and re-use of personal data as open data, and that, conversely, is still compatible with open data requirements.

That, at its core, involves a two-step process. The first step is identifying the points of conflict between open data requirements and data protection secondary legislation. While, intuitively, the idea of balancing open data and data protection would involve evening out competing interests, any modification of the normative status quo allowing the processing of personal data as (proper) open data would entail, in practice, relaxing or eliminating data protection rules and principles. The second step is assessing the lato sensu proportionality (i.e. the suitability, necessity, and proportionality stricto sensu) of a hypothetical measure erasing those points of conflict in favour of whichever right or freedom can be portrayed as being advanced by open data sharing and re-use. Indeed, despite historical and jurisdictional divergences, in the European fundamental rights framework, ‘balancing’ is performed through the lato sensu proportionality assessment of a measure interfering with a right. Such an assessment, in turn, means testing the measure’s suitability for the achievement of the objectives at hand, its (strict) necessity in a democratic society, and the stricto sensu proportionality of the compression to a right vis-à-vis the benefits to a competing right.

There is little doubt that publishing information as open data is suitable to achieve the end goals of a broad array of rights and freedoms, such as freedom of expression and of

\textsuperscript{1563} See Open Data Institute (n 638).

\textsuperscript{1564} For the framework used in this thesis, see generally Alexy (n 931); Alexy (n 981).
information, the right of public access to documents, and freedom of business. Open data is the most permissive data sharing and re-use policy available, perhaps even more than the release of information in the public domain, and its abstract suitability to achieve transparency and accountability, public-sector efficiency, and to foster innovation and growth is not in question. As far as the suitability test goes, and in consideration of the broad range of objectives that can be made to fit the smart city narrative, it seems hard to argue against open data’s theoretical capacity to achieve whichever goal is set by the polity of reference, provided that such a goal can be achieved through data sharing and re-use.

It is the necessity of a measure curtailing data protection to foster open data, and its proportionality strictu senso, that highlight how the idea of balancing data protection and open data is not really about balancing at all. As for the necessity requirement, it does not appear possible to make the case for the strict necessity of such a measure in a democratic society. That threshold cannot surely be met just through a general reference to the promises of the smart city; moreover, on account of the permissive nature of data protection, it cannot be justified by the alternative being ‘closed’ data, either: there are many choices (i.e. “less restrictive measures”, in fundamental rights’ jargon) other than open data sharing, and many of them allow the sharing and re-use of personal data.

As for proportionality in a strict sense, the requirements that must be met to qualify information as open data, by definition, reject the kind of mechanisms that ensure proportionality in personal data processing. Data protection law channels the activity of personal data processing (inter alia) by constraining its accessibility and usability, and information whose accessibility and usability are in any way constrained are not open data in the true sense of the term. In the simplest terms, open data and personal data are meant to be mutually exclusive, an outcome which is reached both by subsuming their interaction under the applicable norms, and by attempting the performance a lato sensu proportionality assessment. Even if the law were not clear on the matter – and it is – it would be impossible to ‘balance’ open data and personal data protection without watering down the meaning of the former, or violating the latter.

The bad news is thus that open data and personal data protection cannot be truly balanced, and indeed speaking of balancing from a normative perspective is not a proper framing altogether. The good news is that it does not really matter all that much: there are few kinds of personal data processing that are entirely forbidden by data protection law – it is mostly a matter of finding a suitable processing architecture, and of accepting the inevitable trade-offs that come with actual balancing. The issue derives from seeing open data as a value to be balanced with data protection, rather than a means through which to foster rights and freedoms. Whenever those rights and freedoms conflict with the right to personal data protection, and there is the need to perform a balancing exercise, open data cannot be a proportionate (and hence reasonable) outcome by its own definition.

Open data exploits IP over information to create a set of licensing conditions that, instead of protecting the IP underlying the licenses, renders it accessible and usable by anybody and for

---

1565 ECHR, Art. 10; Charter, Art. 11.
1566 Charter, Art. 42
1567 Charter, Art. 16.
1568 The machine-readability and the interoperability requirement, it may be argued, enhance the information reusability and ‘findability’, while its availability, usability, and accessibility are on par with the public domain.
1569 I.e. degrees of access control and use restrictions, ensured through both legal and technical means.
any purpose. Akin, for instance, to creative commons, open data leverages freedom of contract (private law) so that the rightsholder – the licensor – can automatically guarantee to the licensee unfettered freedom. Open data does not, and was never meant to, take away public law obligations, like data protection’s, but just to leverage contractual freedom to open up the re-use of information. In such cases, there is no need to resort to balancing and proportionality, as there are no rights that are compressed by countervailing rights and interests: there is no interference upon the (absolute) right to intellectual property, just the exploitation of contractual freedom to create (relative) licensing rights. That is not (entirely) possible vis-à-vis personal data protection, as not all the obligations it mandates and the rights it grants can be waived by the relevant rightsholders. Amongst the rules and principles that cannot be waived by data subjects, there are many that are in insurmountable conflict with open data requirements. Arguing for the necessity to ‘balance’ open data and data protection, rather than setting up a personal data processing architecture that is as permissive as legally and factually possible, would be tantamount to claiming that open data needs to be ‘balanced’ with (intellectual) property rights, rather than setting up an appropriate licensing framework through contractual means. In other words, there is no ‘balancing’ involved between intellectual property and open data, just the application of freedom of contract to the creation of an ‘open’ licensing framework; the same result cannot be achieved within the boundaries of EU data protection law, but that does not mean that there is the need for any actual balancing to be done.

7.5 OPEN DATA AND THE MATERIAL SCOPE OF EU DATA PROTECTION LAW

Open data and data protection are mean to run on parallel tracks, so to speak. The issue is thus determining what is personal and what is not, so that data controllers can benefit from a sufficient amount of legal certainty, and can decide the best data sharing architecture according to the data in consideration. The notion of personal data is however meant to be elastic by design, so that it can provide flexible and contextual protection to the data subjects it relates to. Such elasticity has also drawbacks: the boundaries of the concept of personal data may be somewhat blurry at times, shifting along with the evolution of the state of the art in technology and know-how. The idea of balancing open data and data protection that informs this thesis moves from the postulate that the notion of personal data is undergoing an unrestrained extension, by virtue of the expansive interpretation of the identifiability and the ‘relating to’ requirements, and that such inflation is eating away at the possibility to publish and re-use information as open data.

Regardless of whether that is a matter of concern or not, as the possibility to publish and use information is not ruled out by information being personal, and open data is not the only data sharing architecture conceivable, it is undeniable that the scope of data protection law, both de iure and de facto, has been undergoing a noticeable expansion. That seems, however, largely a consequence of the exponential rise in importance that information, and its

---

1570 See https://creativecommons.org/.
1572 In the sense of having an effect erga omnes.
1573 In the sense of having an effect only inter partes.
1574 Amongst many, see Purtova, ‘The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law’ (n 903); Brkan, ‘The Unstoppable Expansion of the EU Fundamental Right to Data Protection: Little Shop of Horrors?’ (n 903); van Loenen, Kulk and Ploeger (n 977); Aldhouse (n 1296).
processing, has assumed in contemporary society.\textsuperscript{1575} The kinds of information that relate to natural persons, either by virtue of their content or of how they are used, have been constantly increasing over the years, and so have the means and know-how necessary to identify those persons. It is only natural for the material scope of data protection law to follow suit. It is true that the identifiability requirement is interpreted extensively,\textsuperscript{1576} so that even information that relates to a natural person that is indirectly identifiable by someone other than the data controller may be qualified as personal data. It is also true that the ‘relating to’ condition can be satisfied by the existence of many different kinds of relationships between the information and the data subjects: beyond the information whose content relates to a person, even data whose purpose is about a person, or whose processing’s result may have an impact upon a person, may be qualified as personal data.\textsuperscript{1577} In theory, any kind of information may, in the right circumstances, become personal data.

Both requirements must be satisfied at the same time for the information in consideration to be personal data. When one stops considering them in isolation, but rather focuses on their interaction, it results clear that information that relates to the data subject by virtue of its content will always be personal, regardless of its purpose or of the result of its processing. Conversely, information that has a feeble relationship with the data subject (i.e. that relates to him or her by virtue of its purpose or of the result of its processing) must always be supported by auxiliary, identifying data that ties them together to be considered as personal data under EU data protection law. In other words, identifiability functions as a brake to the breadth of the ‘relating to’ requirement, differentiating data that is merely relatable to a natural person from information actually relating to him or her. From this point of view, doctrinal concerns about the material scope of data protection law encroaching upon datasets traditionally deemed ‘safe’ for release as open data (such as many kinds of geographical\textsuperscript{1578} or environmental\textsuperscript{1579} data) seem to be vastly overstated.

Granted, however, anonymity is seldom a certainty, and there have been several cases\textsuperscript{1580} where anonymisation processes have failed in delivering the level of confidentiality they aimed at providing. The truth is that anonymity, just as security, is a status that is not absolute, but rather often dependent on the resources available to a motivated attacker (or intruder). Data protection law, on the other hand, is still based on a neat dichotomy between identifiability and anonymity, which does not really match the current state of the art, and that is largely a legacy of conceptual frameworks developed at its beginning, more than half a century ago. In

\begin{enumerate}
\item \textsuperscript{1575} See e.g. Kitchin, \textit{The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences} (n 4); Schwab (n 817).
\item \textsuperscript{1576} See e.g. CJEU, Patrick Breyer v Bundesrepublik Deutschland (n 845); see also Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40); Article 29 Data Protection Working Party, ‘Opinion 05/2014 on Anonymisation Techniques WP216’ (n 521).
\item \textsuperscript{1577} Article 29 Data Protection Working Party, ‘Opinion 4/2007 on the Concept of Personal Data WP136’ (n 40).
\item \textsuperscript{1578} See van Loenen, Kulk and Ploeger (n 977). That does not mean, of course, that geographical data cannot be personal, both in EU data protection law and in other jurisdictions (see e.g. Teresa Scassa, ‘Geographical Information as “Personal Information”’ [2010] 10 Oxford University Commonwealth Law Journal 185). Cadastral parcels, for instance, may very well be personal data, as they overlay administrative information (see Berlee [n 1208]).
\item \textsuperscript{1579} See Purtova, ‘The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law’ (n 903).
\item \textsuperscript{1580} See the cases cited in Ohm (n 42); Narayanan and Shmatikov (n 1298); Narayanan and Felten (n 1296). See also e.g. Rocher, Hendrickx and de Montjoye (n 1298); Culnane, Rubinstein and Teague (n 1298); de Montjoye and others (n 621).
\end{enumerate}
this respect, this thesis argued two main points. The first is that it would be desirable, from a policy perspective, to align the construction of anonymity and identifiability in data protection law with information security, making it more process-oriented, contextual, and risk-tolerant than it currently is. The second point, which partly follows from the first, is that – from a more practical perspective – there already is an entire discipline (Statistical Disclosure Control (SDC) in Europe, or Statistical Disclosure Limitation (SDL) in the US) whose sole purpose is to ensure a certain trade-off between the data’s utility and its confidentiality guarantees. In fringe situations, and in particularly complex data processing instances, the notion of personal data should not be interpreted as requiring a degree of certainty that is practically unattainable, and should rather focus on the processes enacted by data controllers. In other words, on one hand, all that can reasonably be asked to data controllers dealing with anonymised data is to adhere to SDC/SDL processes and best practices.

On the other hand, computing, know-how, and the “data environment” evolved to a point where de-identifying data by removing direct and indirect identifiers simply does not suffice anymore, and where the decision of releasing information on the internet for anyone to use for whichever purpose they deem fitting should be extensively scrutinised beforehand. It is not the concept of personal data that has changed: the world we live in did. The amount of information that is routinely linked to individuals is much greater than it was when data protection law came to be, and identifying individual records within a de-identified dataset has become much easier; data storage is cheaper and cheaper, and the incentives to its analysis are higher than ever. Anonymisation, on its own, does not guarantee the same level of protection it used to ensure before the rise of the information and communication society, and should now be coupled with other kinds of controls, such as access control and use restrictions, and go beyond the so-called “release-and-forget” data sharing policies. Yet, open data – actual open data – is largely incompatible with technical and legal forms of data protection other than anonymisation. It is a logical conundrum: anonymisation should be coupled with additional controls, and yet whatever restricts the availability and usability of the information contradicts open data’s requirements. To the extent that open data and “release and forget” anonymisation coincide, it would not be unexpected to see a contraction in the release of anonymised information as open data in the coming future. Regardless of personal data and the material scope of EU data protection law, anonymisation alone is just not enough anymore.

There is, nonetheless, a countervailing trend that may mitigate, and perhaps even reverse, the compression that the limits of release-and-forget anonymisation may have on open data sharing and re-use. As testified by the relative provisions in the recast of the PSI Directive, open data, for several reasons, is at least partly shifting from “release-and-forget” data publishing towards interactive, query-based methods. Some implementation of differential

---

1581 See Rubinstein and Hartzog (n 1296).
1582 See generally Hundepool and others (n 1491); Willenborg and De Waal (n 1500); Duncan, Elliot and Salazar-Gonzálezn (n 1500).
1583 See generally the work of Mackey and Elliot, e.g. Mackey and Elliot (n 1433); Elliot and others (n 1374); Elliot and Mackey (n 1434).
1584 See e.g. Gellman (n 1490).
1585 Rectius, the release of an anonymised dataset \( A_1 \) deriving from a (personal) dataset \( A \).
1586 2019 PSI Directive, Art. 5(5).
1587 Not only privacy and data protection – other reasons are, for instance, avoiding duplicated datasets, old data, or inaccurate information.
privacy, for instance, might provide a higher level of protection than “release-and-forget” open data sharing, while still satisfying the requirements on which open data is based, albeit at the cost of a certain amount of utility. Yet again, that is a perspective that construes open data as an end, rather than a means: there are no particular reasons to favour open data over other data sharing and re-use policies when the latter would provide a higher amount of utility, although at the cost of accepting constraints on its accessibility or usability.

7.6 Limitations

The validity of the conclusions of this thesis must be assessed against its limitations, which I am bound to be the first to admit. The pursuit of a balance between open data and data protection, whether for the benefits of the ‘smart city’ or not, can be seen either as a wicked problem or as an impossible question. As for the former, it can be understood as the pursuit of the unattainable balance between an unrestrained and serendipitous data re-use, on the one hand, and the protection guaranteed by a set of rules and principles that constrain its usability and accessibility, on the other. As for the latter, the question “how do we balance open data and data protection?” entails the idea of balancing the fundamental right to personal data protection, and the secondary framework that expresses it in (public) EU law, against a set of licensing conditions based on largely incompatible principles, which is outside of the realm of the legally possible for a plethora of reasons. The framing of this thesis’ research question is thus central in understanding its answers; a different reading of the elements it is made of could have led to partly different results.

A first disclaimer concerns this thesis’ reading of the notion of ‘smart city’. There is no doubt that it may mean something different in other contexts and to other disciplines, but in this thesis, it specifically amounts to the reason to consider the idea of ‘balancing’ open data and data protection in the first place. I understand the reasons to shorthand complex ideas, such as “the instrumentation and datafication of the built environment” or “the interaction between new technologies and urbanism”, but I grew extremely sceptical about the notion of smart city as a reason for (de)regulating personal data processing.

Starting from the assumption that regulatory change ought to move from specific policy objectives, this research began by attempting at defining what a smart city is, what are its requirements, and in which ways would the conflict between open data and data protection hamper the achievement of those requirements. It turned out that there are hundreds of smart city definitions, each more generic than the next, from which it is impossible to enucleate general requirements. Standards containing the wording ‘smart city’ allow for higher precision and comprise specific requirements, but relate to a wide array of different subjects, like urban computing and networking infrastructures, and spatial or environmental data formats, which do not interact with data protection at all or involve information that is not (and should not be) released as open data. This thesis’ reading of the notion of smart city, in other words, has been conditioned by its framing as a justification for pushing forth the idea of balancing open data and data protection, and should be read only within that context.

I do not mean to downplay important disciplines and themes that are generally connected with

---

1588 See generally Wood and others (n 1454).
1589 I.e. contractual terms, based on private law, which have effect only inter partes.
1590 See e.g. the ones listed in Lea (n 208).
1591 I.e. data under the scope of the ePrivacy Directive.
the smart city imaginary: data-driven urbanism, or evidence-based decision-making by local government. When the discourse moves from a technological underpinning, however, I believe that everything except from the technologies considered, the values underlying their development and deployment, and the consequences of the latter, are part of the narrative, and narratives should not drive policymaking.

A second point relates to open data. This thesis assumed a very strict reading of open data means, and relaxing the requirements underlying its definition would undeniably lead to very different results. It has already been noted\textsuperscript{1592} how open data is often used to describe data sharing policies and licensing terms that do not provide \textit{unfettered} accessibility and usability, for instance by foreseeing contractual conditions containing use restrictions. It has also been underlined,\textsuperscript{1593} on the other hand, the importance of maintaining a precise definition of what open data is, so that we may be able to have a common vocabulary and to avoid “open-washing”. From this thesis’ perspective, the need to assume a strict and precise definition of open data derives from the fact that doing otherwise would have signified approaching an untenable research question. If open data were construed as permitting some form of access control and use restriction, there would be no reason not to tune those controls and restrictions for compliance with EU data protection law, and thus no reason for this thesis. The same may be said for the conflation between open data and PSI. If there were identity between PSI and open data, the interaction between open data and data protection would not necessitate of any sort of ‘balancing’: the correct operations to be performed would be the subsumption of a case under a clear set of rules.\textsuperscript{1594}

A third limitation concerns the balancing of competing rights and freedoms. This thesis’ research question adopts the word ‘balancing’, and yet this thesis uses it mainly as a proxy to discuss proportionality. The reason for this framing is, beyond my own personal reading, of a disciplinary, methodological, and geographical nature. It is a matter of discipline as, despite being interdisciplinary in aspiration, this is still a dissertation in law. I understand ‘balancing’ would mean something different to political scientists or policymakers, for instance, but I do not belong to those categories, and I see balancing from a legal perspective, and not from a political one. The methodological aspect is also relevant: this thesis frames proportionality as the method through which balancing is performed, and I move from the assumption that there needs to be a framework through which power’s discretion can be curtailed, and its processes can be formalised. Finally, from a geographical and jurisdictional standpoint, this dissertation moves from an EU perspective; the balancing of competing rights and freedoms would assume different characteristics in other jurisdictions, for instance the US.

7.7 FURTHER RESEARCH

Investigating a topic has often the side effect of unveiling several other promising areas to explore. In this thesis’ case in particular, it is especially important to outline the areas in which, I believe, further research would be beneficial to the understanding of the interaction between the law, information and communication technologies, and their effect on society. This

\textsuperscript{1592} “Even many presumptively “open” data sets require assent to terms of use agreements”: Rubinstein and Hartzog (n 1296) 37.

\textsuperscript{1593} Pollock (n 635).

\textsuperscript{1594} E.g. the exception on grounds of data protection ex Art. 1(2)(h) of the 2019 PSI Directive, and the non-prejudice clause ex Art. 1(4).
dissertation’s main outcome has been an opposition to a narrative that would lead to needless deregulation; its secondary result has been highlighting where the actual issues lie, and providing some interpretative tools to nuance their discussion.

A first promising research area concerns the relationship between data protection law and statistical disclosure control/limitation, both from a technical and practical perspective and from a normative one. In general, the alignment between data protection law and statistics is an underexplored area. It would seem that, aside from the obvious overlap in scope between statistical disclosure control personal data protection, as they both involve granting a certain amount of confidentiality to natural persons (or statistical units), SDC might provide the means for building tolerance for genuine uncertainty and risk within the data protection framework. An useful exercise could be mapping inputs and outputs of SDC methods to their equivalents under data protection law, so that data controllers may benefit from additional legal certainty in determining, in fringe situations, which data is personal, which is anonymous, and which is pseudonymous.

Beyond the smart city narrative, the instrumentation and datafication of the built environment presents significant problematics that should, in my opinion, be tackled by starting from the fundamentals. What are the specific technologies that are being deployed within the urban environment, what are the values driving their development, and what are the consequences that derive from it? I fear that the smart city may be used as a trojan horse, a narrative that dissimulates an eminently neoliberal leaning through corporate storytelling, and a push towards the securitisation of the built environment.

A third area of further research concerns the cost/benefit analysis and the risk assessment of open data vis-à-vis other options in the openness spectrum, in particular when anonymised data are concerned. To what extent do technical and contractual access controls and use restrictions lower the benefits that may derive from open data sharing and re-use? When the implementation of technical and legal controls allows compliance with data protection law, it may be inefficient to anonymise the data, degrading its utility by meeting the high threshold required by EU data protection law.

7.8 CONCLUSION

In 2015, I was tasked with the development of a construct to balance open data and data protection in the context of the smart city environment. This research began by investigating what a smart city is, and how open data and data protection interact, and ended up questioning the necessity of the pursuit of a balance between open data and data protection, its feasibility, and the general framing of the research question in the first place. Arguing that data protection should be somehow balanced with open data, be it in the smart city or not, is a misleading framing: it portrays open data as a value or a goal on its own, rather than a means that can be used to uphold several different rights and values.

Once the meaning of proportionality is fleshed out, it results evident that the idea of balancing open data and data protection does not only fail a proportionality test, but cannot be run

---

1595 I.e. in the blurry boundaries of the material scope of EU data protection law.
1596 See Rubinstein and Hartzog (n 1296).
1597 As e.g. in Hu and others (n 1332).
1598 See Mourby and others (n 1435).
1599 See Wu (n 1296).
through it at all. The reason is that open data’s role within a proportionality assessment is not as an input, but as a potential output or result. The matter is that open data is at the most permissive end of the openness spectrum, and – as long as more proportionate options are available – it does not seem possible to configure a situation where, when different rights compete, open data is a necessary or (lato sensu) proportionate outcome.

That is, however, to a large extent either uncontroversial or unimportant. The dreaded overextension of the concept of personal data seems to be, to a closer examination, merely a response to the application of a concept that is as old as data protection law is to the evolving technological landscape. Furthermore, data protection law is a permissive regime by nature: few kinds of processing are actually forbidden – too few, arguably – none of which is necessary for smart cities to run. The conflict between the principles on which open data is based and personal data protection law makes it so that, even when freely available on the internet, personal data can never be truly open data. That does not however seem a problem per se, as the range of options is not limited to open and ‘closed’ data, but encompasses many degrees of the openness spectrum. Personal data is not ‘closed’ by default, and compliance with data protection law does not necessarily mean that the information cannot be made accessible or usable to third parties.

Should a balance between open data and the right to personal data protection be pursued? I believe it should not, as it cannot be reached. The balance between data protection and the rights and freedoms underlying open data is an entirely different thing.
BIBLIOGRAPHY

Ahonen P and others, Safeguards in a World of Ambient Intelligence, vol 1 (David Wright and others eds, Springer Science & Business Media 2008)

Ajuntament de Barcelona, ‘Pla Digital de l’ Ajuntament de Barcelona Mesura de Gestió Ètica i Responsable de Dades: Barcelona Data Commons’

Alawadhi S and others, ‘Building Understanding of Smart City Initiatives’, Electronic government (Springer 2012)


Albrecht JP, Hands off Our Data! (The Greens | EPA 2015)


— —, A Theory of Constitutional Rights (Oxford University Press 2009)


Angelidou M, ‘Smart City Policies: A Spatial Approach’ (2014) 41 Cities S3

— —, ‘Smart Cities: A Conjecture of Four Forces’ [2015] Cities 95


— —, ‘Opinion 03/2013 on Purpose Limitation’ (2013)

— —, ‘Opinion 06/2013 on Open Data and Public Sector Information (“PSI”) Reuse WP207’ (2013)

— —, ‘Opinion 05/2014 on Anonymisation Techniques WP216’ (2014)


— —, ‘Statement 14/EN WP 218 on the Role of a Risk-Based Approach in Data Protection Legal Frameworks’ (2014)


Balboni P and others, ‘Legitimate Interest of the Data Controller New Data Protection Paradigm: Legitimacy
Grounded on Appropriate Protection’ (2013) 3 International Data Privacy Law 244
Bamberger KA and Mulligan DK, Privacy on the Ground: Driving Corporate Behavior in the United States and Europe (MIT Press 2015)
Barry E and Bannister F, ‘Barriers to Open Data Release: A View from the Top’ (2014) 19 Information Polity 129
Batty M and others, ‘Smart Cities of the Future’ (2012) 214 The European Physical Journal Special Topics 481
Bradshaw T, ‘Mobiles Could Be the Secret to “Smart” Cities’ Financial Times (22 February 2016)
Bretthauer D, ‘Open Source Software: A History’
———, ‘The Concept of Essence of Fundamental Rights in the EU Legal Order: Peeling the Onion to Its Core’ [2018] European Constitutional Law Review 1
BSI and Imperial College London Consultants, ‘Mapping Smart City Standards Based on a Data Flow Model’ (2015)
Bygrave LA, ‘Data Protection by Design and by Default: Deciphering the EU’s Legislative Requirements’ (2017) 4 Oslo Law Review 105
Calzada I, ‘(Smart) Citizens from Data Providers to Decision-Makers? The Case Study of Barcelona’
Capdevila I and Zarlenga MI, ‘Smart City or Smart Citizens? The Barcelona Case’ (2015) 8 Journal of Strategy and Management 266
Caragliu A, Del Bo C and Nijkamp P, ‘Smart Cities in Europe’ (2011) 18 Journal of urban technology 65
Carrara W and others, Creating Value through Open Data: Study on the Impact of Re-Use of Public Data Resources (European Commission 2015)
———, ‘Open Data Goldbook for Data Managers and Data Holders’ (2016)
———, ‘Privacy by Design: Leadership, Methods, and Results’ in Serge Gutwirth and others (eds), European Data Protection: Coming of Age (Springer Netherlands 2013)
Christl W and Spiekermann S, Networks of Control (Facultas 2016)
———, ‘Legal Research’ in Andrew Knight and Les Ruddock (eds), Advanced research methods in the built environment (John Wiley & Sons 2009)
Clifford D and Ausloos J, ‘Data Protection and the Role of Fairness’ [2018] Yearbook of European Law
Cocchia A, ‘Smart and Digital City: A Systematic Literature Review’ in Renata Paola Dameri and Camille Rosenthal-Sabroux (eds), Smart City (Springer 2014)

Cole PE, ‘New Challenges to the U.S. Multinational Corporation in the European Economic Community: Data Protection Laws’ 893


Committee on Data Protection chairman Sir Norman Lindop, ‘Report of the Committee on Data Protection’ (Norman Lindop ed, HMSO 1978)


Craig P, ‘Development of the EU’ in Catherine Barnard and Steve Peers (eds), European Union Law (Oxford University Press 2014)


Cuijpers C and Koops BJ, ‘Smart Metering and Privacy in Europe: Lessons from the Dutch Case’ in Serge Gutwirth and others (eds), European Data Protection: Coming of Age (Springer 2013)


Culnane C, Rubinstein BIP and Teague V, ‘Health Data in an Open World’ (2017)


Dalerius T, ‘Towards a Methodology for Statistical Disclosure Control’ (1977) 15 Statistik Tidskrift 1


Danezis G and others, ‘Privacy and Data Protection by Design - from Policy to Engineering’ (European Union Agency for Network and Information Security 2014)
Davies SG, ‘Re-Engineering the Right to Privacy: How Privacy Has Been Transformed from a Right to a Commodity’ in Marc Rotenberg and Philip E Agre (eds), Technology and Privacy: the New Landscape (MIT Press 1997)


— — , ‘Legal Safeguards for Privacy and Data Protection in Ambient Intelligence’ (2009) 13 Personal and ubiquitous computing 435


de Montjoye Y-A and others, ‘Unique in the Crowd: The Privacy Bounds of Human Mobility’ (2013) 3 Scientific Reports 1376


De Vries M and others, ‘POPSIS – Pricing Of Public Sector Information Study’ (2011)


DeLong JB and Froomkin AM, ‘Speculative Microeconomics for Tomorrow’s Economy’ (2000) 5 First Monday


Docksey C, ‘Articles 7 and 8 of the EU Charter: Two Distinct Fundamental Rights’ in Alain Grosjean (ed), Enjeux européens et mondiaux de la protection des données personnelles (Éd Larcier 2015)
Doctorow C, *Information Doesn’t Want to Be Free: Laws for the Internet Age* (McSweeney’s 2014)


——, ‘LAPS! Policy Recommendation n. 4 Privacy and Personal Data Protection’


Dufourmont H and others, ‘Extended Impact Assessment of INSPIRE Based on Revised Scope’ (2004)


European Commission, ‘INSIPRE – History’


——, ‘Memorandum of Understanding between Commissioners Wallströrm, Solbes, Busquin Infrastructure for Spatial Information in Europe’ (European Commission 2002)

——, ‘Communication from the Commission to the European Parliament, the Council, the Economic and Social


———, ‘Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and The Committee of The Regions “Open Data An Engine for Innovation, Growth and Transparent Governance” COM(2011) 882 Final’ (2011)


———, ‘Public Sector Information Group’ (2013)


———, ‘Open Data Incubator for Europe (ODINE)’ (2016)

———, ‘Rolling Plan for ICT Standardisation’ (2016)

———, ‘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” (SWD(2017) 2 Final’ (2017)

———, ‘Synopsis Report Consultation on the “Building a European Data Economy” Initiative’ (2017)


European Data Protection Supervisor, ‘Assessing the Necessity of Measures That Limit the Fundamental Right to the Protection of Personal Data: A Toolkit’


Fabbrini F, Fundamental Rights in Europe: Challenges and Transformations in Comparative Perspective (Oxford University Press 2014)

Farnsworth DP, ‘Data Privacy or Data Protection and Transborder or Transnational Data Flow, an American’s View of European Legislation’ (1983) 11 International Business Lawyer 114

Fernandez M, ‘Smart Cities of the Future?’ in Drew Hemment and Anthony Townsend (eds), Smart citizens (FutureEverything Publications 2013)


Finn RL, Wright D and Friedewald M, ‘Seven Types of Privacy’ in Serge Gutwirth and others (eds), European Data Protection: Coming of Age (Springer 2013)

Convention 108’ (2012) 2 International Data Privacy Law 68


Group of 8, ‘G8 Open Data Charter’

GSMA, ‘Keys to the Smart City – How Mobile Operators Are Playing a Crucial Role in the Development of Smart Cities’ (2016)


Harbinja E, ‘Does the EU Data Protection Regime Protect Post-Mortem Privacy and What Could Be the Potential Alternatives’ (2013) 10 SCRIPTed 19


———, ‘The Case for Online Obscurity’ (2013) 101 Cal. L. Rev. 1

Hemment D, Townsend A and et al., Smart Citizens (Drew Hemment and Anthony Townsend eds, FutureEverything Publications 2013)


Hielkema H and Hongisto P, ‘Developing the Helsinki Smart City: The Role of Competitions for Open Data Applications’ (2013) 4 Journal of the Knowledge Economy 190

Hijmans H, The European Union as Guardian of Internet Privacy: The Story of Art 16 TFEU (Springer 2016)


Hildebrandt M, Smart Technologies and the End (s) of Law: Novel Entanglements of Law and Technology (Edward Elgar Publishing 2015)


Hollands RG, ‘Will the Real Smart City Please Stand up? Intelligent, Progressive or Entrepreneurial?’ (2008) 12 City


Hondius FW, Emerging Data Protection in Europe (North-Holland Publishing Company 1975)


A Decade of International Data Protection’ (1983) 30 Netherlands International Law Review 103


Hu R and others, ‘Bridging Policy, Regulation, and Practice? A Techno-Legal Analysis of Three Types of Data in the GDPR’ in Ronald Leenes and others (eds), Computers, Privacy and Data Protection: the Age of Intelligent Machines (Hart Publishing 2017)

Hundepool A and others, Statistical Disclosure Control (John Wiley & Sons 2012)


Johnson PA and others, ‘The Cost(s) of Geospatial Open Data’ (2017) 21 Transactions in GIS 434


Keymolen E and Voorwinden A, ‘Can We Negotiate? Trust and the Rule of Law in the Smart City Paradigm’ [2019] International Review of Law, Computers & Technology 1


——, The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences (Sage 2014)

——, ‘The Real-Time City? Big Data and Smart Urbanism’ (2014) 79 Geojournal 1

——, ‘Big Data, New Epistemologies and Paradigm Shifts’ (2014) 1 Big Data & Society

——, ‘Data-Driven, Networked Urbanism’ (2015) 14


Klauser F, ‘Through Foucault to a Political Geography of Mediation in the Information Age’ (2013) 68 Geographica Helvetica 95


Koops BJ and others, ‘A Typology of Privacy’ (2016) 38 University of Pennsylvania Journal of International Law

Korff D and Shadbolt N, ‘Public Information: Cause for Celebration or Concern’ [2010] Public and Science 10

Kosta E, Consent in European Data Protection Law (Martinus Nijhoff Publishers 2013)

Kroes N, ‘From Crisis of Trust to Open Governing’ (2012)
——, ‘The Big Data Revolution’ (2013)
——, Transborder Data Flows and Data Privacy Law (Oxford University Press Oxford 2013)
Lauriault T and others, ‘Open Smart Cities in Canada: Environmental Scan and Case Studies’ (2018)
Lefebvre H, Le Droit à La Ville (Paris Anthropos 1968)
Lessig L, Code and Other Laws of Cyberspace (Basic books 1999)
Lindahl L, ‘On Robert Alexy’s Weight Formula for Weighing and Balancing’ in Augusto Silva Dias (ed), Liber Amicorum de José de Sousa e Brito (Edicoes Almedina 2009)
Lynskey O, ‘From Market-Making Tool to Fundamental Right: The Role of the Court of Justice in Data Protection’s Identity Crisis’ in Serge Gutwirth and others (eds), European Data Protection: Coming of Age (Springer Netherlands 2013)
——, ‘Deconstructing Data Protection: The “Added-Value” of a Right to Data Protection in the EU Legal Order’ (2014) 63 International and Comparative Law Quarterly 569
——, The Foundations of EU Data Protection Law (Oxford University Press 2015)

——, ‘From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era’ in Linnet Taylor, Luciano Floridi and Bart van der Sloot (eds), Group Privacy (Springer 2017)


Martin S and others, ‘Risk Analysis to Overcome Barriers to Open Data’ (2013) 11 Electronic Journal of e-Government 348


McBratney AB and Jarrett LE, ‘Securitisation’ in Damien J Field, Cristine LS Morgan and Alex B McBratney (eds), Global Soil Security (Springer International Publishing 2017)


Meessen P and others, ‘Decentralised Models for Data and Identity Management: Blockchain and ABC MVPs’ (2018) DECODE D3.8

Meijer A and Bolivar MPR, ‘Governing the Smart City: A Review of the Literature on Smart Urban Governance’ [2015] International Review of Administrative Sciences

Meldman JA, ‘Centralized Information Systems and the Legal Right to Privacy’ (1968) 52 Marq. L. Rev. 335


Meyerson B, ‘Smartphones May Enable Smart Cities’ Financial Times (4 March 2013)

Millard C, Cloud Computing Law (Oxford University Press 2013)


——, The Assault on Privacy: Computers, Data Banks, and Dossiers (University of Michigan Press 1971)

——, ‘The Dossier Society’ [1971] U. Ill. LF 154


Rid T, Rise of the Machines. A Cybernetic History (W W Norton 2016)


Rossi L and e Silva PV, ‘The Normative Development of Access to Documents’ in Leonor Rossi and Patricia Vinagre e Silva (eds), Public Access to Documents in the EU (Bloomsbury Publishing 2017)


Rubinstein IS, ‘Regulating Privacy by Design’ (2011) 26 Berkeley Tech. LJ 1409


Schneier B, Liars and Outliers: Enabling the Trust That Society Needs to Thrive (John Wiley & Sons 2012)


Secretary’s Advisory Committee and on Automated Personal Data Systems, ‘Records, Computers, and the Rights of Citizens’ (1973)

Selbst AD and Powles J, ‘Meaningful Information and the Right to Explanation’ (2017) 7 International Data Privacy Law 233


Shepard M and Simeti A, ‘What’s So Smart About the Smart Citizen’ in Drew Emment and Anthony Townsend (eds), Smart Citizens (FutureEverything Publications 2013)


Söderström O, Paasche T and Klauser F, ‘Smart Cities as Corporate Storytelling’ (2014) 18 City

Soffel J, ‘Rio’s “big Brother” Control Room Watches over the City’ CNN (29 August 2013)


Stoycheff E, ‘Under Surveillance: Examining Facebook’s Spiral of Silence Effects in the Wake of NSA Internet Monitoring’ (2016) 1 Journalism & Mass Communication Quarterly 16


Sweeney L, Abu A and Winn J, ‘Identifying Participants in the Personal Genome Project by Name’ (2013)
Sweeney L and Yoo JS, ‘De-Anonymizing South Korean Resident Registration Numbers Shared in Prescription Data’ [2015] Technology Science
Sweet AS and Mathews J, ‘Proportionality Balancing and Global Constitutionalism’ 72
Tannen D, Hamilton HE and Schiffrin D, The Handbook of Discourse Analysis (Wiley Online Library 2015)
Taylor L, Floridi L and van der Sloot B, Group Privacy: New Challenges of Data Technologies, vol 126 (Springer 2016)
Thomas V and others, ‘Where’s Wally? In Search of Citizen Perspectives on the Smart City’ (2016) 8 Sustainability 207
——, Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia (WW Norton & Company 2013)
Tranberg CB, ‘Proportionality and Data Protection in the Case Law of the European Court of Justice’ (2011) 1 International Data Privacy Law 239
Tufekci Z, ‘Engineering the Public: Big Data, Surveillance and Computational Politics’ (2014) 19 First Monday
Tzanou M, ‘Data Protection as a Fundamental Right next to Privacy? “Reconstructing” a Not so New Right’ (2013) 3 International Data Privacy Law 88
Ursic H, ‘Unfolding the New-Born Right to Data Portability: Four Gateways to Data Subject Control’ (2018) 15 SCRIPTed 42
——, ‘Architecture and Value-Sensitive Design’ in Claudia Basta and Stefano Moroni (eds), Ethics, design and planning of the built environment (Springer 2013)
van der Sloot B, ‘Public Sector Information & Data Protection: A Plea for Personal Privacy Settings for the Re-Use of PSI’ (2011) 20 Informatica e Diritto 219


van Loenen B, Developing Geographic Information Infrastructures: The Role of Information Policies (IOS Press 2006)

van Loenen B and Grothe M, ‘INSPIRE Empowers Re-Use of Public Sector Information’ (2014) 9 International Journal of Spatial Data Infrastructures Research 86


van Zoonen L, ‘Privacy Concerns in Smart Cities’ (2016) 33 Government Information Quarterly

Vickery G, ‘Review of Recent Studies on PSI Re-Use and Related Market Developments’ (2011)


— — , Privacy as Trust: Information Privacy for an Information Age (Cambridge University Press 2018)


Warren SD and Brandeis LD, ‘The Right to Privacy’ (1890) 4 Harvard Law Review 193

Washburn D and others, ‘Helping CIOs Understand “Smart City” Initiatives’ (2010)

Waterfront Toronto, ‘Request for Proposals: Innovation and Funding Partner for the Quayside Development Opportunity RFP No.: 2017-13’ (2017)


Weller T, ‘The Information State An Historical Perspective on Surveillance’ in David Lyon, Kirstie Ball and Kevin D Haggerty (eds), Routledge handbook of surveillance studies (Routledge 2012)

Westin AF, Privacy and Freedom (Athenaeum 1967)

Westin AF and Baker MA, Databanks in a Free Society: Computers, Record-Keeping, and Privacy; Report (Crown 1972)

Wiebe A and Dietrich N, Open Data Protection-Study on Legal Barriers to Open Data Sharing-Data Protection and PSI (Universitätsverlag Göttingen 2017)

Wiig A, ‘IBM’s Smart City as Techno-Utopian Policy Mobility’ (2015) 19 City 258

3 Scientific data
Wonderlich J, ‘Ten Principles for Opening up Government Information’ (Sunlight Foundation 2010)
Wu FT, ‘Defining Privacy and Utility in Data Sets’ (2012) 84 University of Colorado Law Review 1117
Yaga D and others, ‘NISTIR 8202 Blockchain Technology Overview’ (National Institute of Standards and Technology 2018)
Yin C and others, ‘A Literature Survey on Smart Cities’ (2015) 58 Science China Information Sciences
Zarsky TZ, ‘Incompatible: The GDPR in the Age of Big Data’ (2016) 47 Seton Hall L. Rev. 995
Zuiderwijk AMG, ‘Open Data Infrastructures: The Design of an Infrastructure to Enhance the Coordination of Open Data Use’ (TU Delft 2015)