

Emerging models of data governance in the age of datafication

Marina Micheli¹ , Marisa Ponti², Max Craglia¹ and Anna Berti Suman³

Big Data & Society
July–December: 1–15
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2053951720948087
journals.sagepub.com/home/bds



Abstract

The article examines four models of data governance emerging in the current platform society. While major attention is currently given to the dominant model of corporate platforms collecting and economically exploiting massive amounts of personal data, other actors, such as small businesses, public bodies and civic society, take also part in data governance. The article sheds light on four models emerging from the practices of these actors: data sharing pools, data cooperatives, public data trusts and personal data sovereignty. We propose a social science-informed conceptualisation of data governance. Drawing from the notion of data infrastructure we identify the models as a function of the stakeholders' roles, their interrelationships, articulations of value, and governance principles. Addressing the politics of data, we considered the actors' competitive struggles for governing data. This conceptualisation brings to the forefront the power relations and multifaceted economic and social interactions within data governance models emerging in an environment mainly dominated by corporate actors. These models highlight that civic society and public bodies are key actors for democratising data governance and redistributing value produced through data. Through the discussion of the models, their underpinning principles and limitations, the article wishes to inform future investigations of socio-technical imaginaries for the governance of data, particularly now that the policy debate around data governance is very active in Europe.

Keywords

Data governance, Big Data, digital platforms, data infrastructure, data politics, data policy

Introduction and context

This article examines and synthesises current discourses and practices on the governance of data. It scrutinises different approaches for accessing, controlling, sharing and using data in today's platform economy and derives four emerging models of data governance. The current platform economy is mainly characterised by the asymmetry of power of a few technology corporations and telecommunication companies that have established de-facto quasi-data monopolies. The negative societal implications of this system, including biases in algorithmic decision-making, nudging and manipulation, and privacy violations are increasingly highlighted by research (e.g., Beer, 2017; Kitchin, 2017; Taylor, 2017), while scandals such as Cambridge Analytica raise the awareness among public opinion and policy makers, at least in Europe, that the distortions of this model need to be addressed. The General Data Protection Regulation (GDPR) is an

important step in this direction, even if with some limitations (Delacroix and Lawrence, 2019), and further new measures are being prepared in the European Union, including a Digital Services Act and a Data Act (European Commission (EC), 2020a). Given these forthcoming legislative changes, we deliberately eschew the legislative dimension in this article and focus instead on the social practices implemented and

¹Joint Research Centre, European Commission, Varese, Italy

²Department of Applied Information Technology, University of Gothenburg, Gothenburg, Sweden

³The Tilburg Institute for Law, Technology, and Society (TILT), Tilburg University, and the Joint Research Centre, European Commission, Varese, Italy

Corresponding author:

Marina Micheli, Joint Research Centre, European Commission, Via Fermi 2749, 21027 Ispra, Varese, Italy.

Email: marina.micheli@ec.europa.eu



theorised for governing (big) data. In doing so we respond to scholars and policy makers advocating for a reorientation of the scientific gaze, from the critique of the current data landscape to the possibilities of agency from the ‘bottom up’ (Beraldo and Milan, 2019; Couldry and Powell, 2014; Kennedy et al., 2015), the ‘good data practices’ (Mann et al., 2019) and the ‘alternative’ data governance models (Morozov and Bria, 2018; Carballa Smichowski, 2019). We contribute to this line of inquiry adopting a *social science-informed* approach to data governance that emphasises power relations between actors. Informed by science and technology studies (STS) and critical data studies (CDS), we used data infrastructure as an analytical lens to conceptualise each data governance model as a situated, contingent and relational instantiation of the stakeholder roles, their interrelationships, their articulations of value, and the organisations of governance principles, instruments and mechanisms in each model. Through the notion of data politics, instead, we emphasised the power (un)balances looking at how power relations and asymmetries affect the processes and the goals of particular governance models and how the value created from data is redistributed in each model.

The article addresses the following questions: (1) What configurations of roles and relationships between stakeholders can we identify in the emerging models of data governance? To what extent are other actors beyond corporate data platforms able to participate? (2) What kind of value is pursued and how is it redistributed across actors and society? What mechanisms and arrangements are set in place to generate value from the data?

The dominant model of data governance in current ‘platform society’ is the one established by a few corporate big tech platforms (Srnicsek, 2017; Van Dijck et al., 2018; Zuboff, 2015), but other actors beyond ‘big tech’ are progressively becoming involved in controlling personal data and producing value from it through different data governance models. These alternative models are the focus of this article. We are cognizant that platforms too have been addressing concerns related to power asymmetries, for instance through technical changes and transparency efforts (Gorwa and Garton Ash, 2020). However, these attempts often function as ‘technologies of legitimization’ (Harrison and Mort, 1998: 60) as they do not enable a *true engagement* with the public, nor significant power shifts. For this reason, we turn our attention away from big corporate platforms and look instead at the practices for data access and control developed by societal actors. These practices are a fertile context for developing socio-technical imaginaries

for data that might influence how (big) data will be governed in the future.

The policy relevance of this article stems from the current geopolitical competition around Artificial Intelligence (AI) seen as central to the development of our increasingly digital societies (Craglia et al., 2018; EC, 2018a). China, the USA, and many other countries are investing heavily in AI, and Europe is responding with its own coordinated plan (EC, 2018b) and a strategy for data (EC, 2020b) that stresses the value of extracting greater benefits from, and exercising greater control over, European data. In this phase of the policy and academic debate, we see the emergence of many terms and concepts like data trusts, data sovereignty, and so on with unclear or contradictory definitions and usage. The analysis of data governance models of this article clarifies some of these concepts and is therefore relevant to both research and policy.

The article is organised as follows: after this initial introduction, the next section defines our conceptualisation of data governance, while the subsequent section details the research strategy and the dimensions that informed the analysis. We then illustrate the data governance models that resulted from our research. In the discussion, we critically examine these models addressing the research questions. After specifying the limitations of the study, we conclude highlighting the contribution of the article to the current policy debate on data governance.

Data governance: A social science-informed definition

The term governance has been extensively used in the last two decades but its meaning is still ambiguous (Colebatch, 2014; Rhodes, 1996). Our understanding is informed by existing debates in the political science and risk scholarship (Colebatch, 2014; Kooiman, 2003; Rhodes, 1996) where, for example, governance has been framed as ‘the multitude of actors and processes that lead to collective binding decisions’ (Van Asselt and Renn, 2011: 431). Governance broadly refers to the web of actors involved, with different roles, in the process of governing a system. The term stresses a discontinuity from so-called ‘command-and-control’ by the State, and acknowledges that a broader set of actors and institutions are (also) involved in managing societies (private sector, civil society and other non-government entities; Kooiman, 2003). Governing is the result of a process, which does not only occur through rule making and rule enforcing but develops also from (social) interactions, cooperation and negotiations between stakeholders at the horizontal level (Colebatch, 2014).

The notion of governance has a descriptive meaning but also has a normative scope capturing a way of ‘rule-making’ (including strictly laws, but also regulations, standards, etc.) and the relative allocation of responsibilities and liabilities. Governance embraces a broader transition from more centralised to decentralised forms of rule-making. Van Asselt and Renn (2011: 435) underline the normative understanding of governance describing it as a ‘model or framework for organizing and managing society’. This normative understanding promotes wider participation and increased accountability in formal decision-making processes. Participation of a broader set of stakeholders itself is indeed acknowledged as a *value* of good governance. From this point of view, the normative understanding of governance operates as a guiding and framing lens for our analysis.

As Wolf (2002) puts it, the governance phenomenon takes place within *horizontally* organised structures where both state and non-state actors (including citizens) interact. Yet this is easier in theory than in practice as power disparities among actors continue to exist and matter. Market actors often benefit from these more fluid allocations of power and responsibilities (DeNardis, 2019; Srnicek, 2017), at the detriment of less (economically) powerful actors such as citizens, communities and civil society organisations (Heeks and Shekhar, 2019; Taylor and Broeders, 2015). This is particularly evident in the context of mediated environments where the dominant position of social media platforms, acting as intermediaries between users and information and infrastructures for (online) participation, leads to a ‘privatization of internet governance’ (DeNardis, 2019).

Based on this understanding of governance, we examine in this article ways in which personal data collected through datafication processes is and could be governed. There is an extensive literature on data governance in the fields of privacy regimes (Bennett and Raab, 2018) and of information systems, from which we draw for the definition of the analytical dimensions (Abraham et al., 2019; Winter and Davidson, 2019; Khatri and Brown, 2010). This contribution, however, adopts a social science-informed perspective of data governance that complements other framings, such as those of platform governance or privacy and data protection law. Our perspective on data governance draws in particular from STS and CDS, which informed our work through concepts of data infrastructure (Kitchin and Laurialt, 2014) and data politics (Bigo et al., 2019; Ruppert et al., 2017).

Following the conceptualisation of infrastructure in STS as a heterogeneous, relational, and complex socio-technical ‘assemblage’ (Slota and Bowker, 2017), a data infrastructure is seen as an evolving ecosystem with a

plurality of actors having multiple interests, agendas, goals and strategies, and interacting with an array of tools, mechanisms, systems, interfaces and devices for governing data (Kitchin and Laurialt, 2014). A data infrastructure is implemented not only to support certain practices, but also to cultivate a specific imaginary, that is, a particular vision of data and its possibilities (Beer, 2017; Gray et al., 2018).

The notion of data politics emphasises the ‘performative power of data’. It understands data not only for its representational capacities, but also as a force ‘generative of new forms of power relations’ (Ruppert et al., 2017: 2), as how data is collected and processed generates power imbalances and information asymmetries in bringing into being the subjects and objects that such data concerns. Data politics conceives data as ‘an object of investment’ that is ‘produced by the competitive struggles of (actors) who claim stakes in its meaning and functioning’ (Ruppert et al., 2017: 5). It underlines the role of data subjects and asks questions about their position *vis a vis* digital platforms. Data politics looks both at how data subjects are governed, and how they can intervene in the data regimes by recognising and claiming their rights and being active in the politics of data with their everyday acts (Kennedy and Moss, 2015; Ruppert et al., 2017).

Drawing on CDS and STS we use data infrastructure as an analytical lens to conceptualise the identified data governance models as a situated, contingent and relational instantiation of the stakeholder roles, their interrelationships, their articulations of value, and the organisations of governance principles, instruments and mechanisms in each model. Likewise, by acknowledging the asymmetries of the current data landscape and the public debate on how to challenge them, we direct our attention to the issue of power within the politics of data. Informed by these scholarships and concepts, we understand data governance as *the power relations between all the actors affected by, or having an effect on, the way data is accessed, controlled, shared and used, the various socio-technical arrangements set in place to generate value from data, and how such value is redistributed between actors*. This implies examining the enacting subjects and objects involved and considering how power relations and asymmetries affect the processes and the goals of particular governance models. In this context, data governance does not refer to data management practices, but to the decisions made over data, who is able to make such decisions and thus to influence the way data is accessed, controlled, used and benefited from (cf. Abraham et al., 2019; Khatri and Brown, 2010; Rosenbaum, 2010; Winter and Davidson, 2019).

Based on this conceptualisation, the article examines four emerging data governance models. These models

could be understood as inventive practices that problematise current arrangements and reassemble them in accordance to the interests of the actors involved. Through the analysis we can scrutinise the ‘desirable futures’ these models promote (Jasanoff, 2015) and see whether they address the asymmetries of power of the current data landscape.

Research strategy

To identify emerging data governance models, we delved into grey and academic literature, as well as news articles and websites of recent projects and initiatives. The collection of resources started in preparation of a workshop held in October 2018 on data governance with 17 invited experts from academia, public sector, policymaking, research and consultancy firms (Micheli et al., 2018). The workshop highlighted a widespread lack of knowledge and practical understanding of alternative models to the ‘data extraction’ approach of big online platforms (Zuboff, 2015), the need to find ways to use data collected by private companies for the public interest, and the urgency to consider data subjects as key stakeholders for the governance of data (Micheli et al., 2018). The outcomes from the workshop were therefore critical in informing our study.

As a preliminary research strategy, we retrieved articles using Google Scholars and the Web of Science Core Collection, but we then decided to proceed without being constrained by keywords since the results were not pertinent¹ to our research objectives and our conceptualisation of data governance. Consequently, we adopted a flexible search strategy and used a snowballing approach including progressively new sources according to their relevance to the theme of interest. The initial sources considered for this research were identified for the preparation of the workshop and from the inputs provided by the workshop participants (Micheli et al., 2018). Such resources addressed the power unbalances of the current datafied society and advocated a democratic and equitable digital transformation, with particular emphasis on socially beneficial uses of data held by the public sector and citizens empowerment through data (e.g. Couldry and Powell, 2014; Andrejevic, 2014; Kennedy and Moss, 2015; Symons and Bass, 2017; Morozov and Bria, 2018; Villani, 2018; Winter and Davidson, 2019; Shkabatur, 2019; Ilves and Osimo, 2019; Carballa Smichowski, 2019). A subsequent step was to review related work, which addressed similar issues or was directly linked to the sources examined. Simultaneously, we kept track of new publications and on-going projects or initiatives. The review

strategy proceeded iteratively, until the typology of the models was consolidated.

We considered the above strategy appropriate for two reasons: the understanding of data governance developed for this contribution diverges from interpretations prevailing in scholarly literatures (coming especially from management, information systems, and law), thus it needs a broader research procedure. Furthermore, the object of the study is a rapidly evolving field for which an established shared vocabulary is lacking. The various labels that are being proposed in current data policy discourses tend to be used equivocally to refer to different concepts (technical solutions, legal frameworks, economic partnerships), with their meaning shifting according to the context. Therefore, we did not bind the research of resources to predefined labels. Although the method is fairly speculative compared to systematic literature reviews, it was nevertheless suitable for the purpose of synthesising, and critically inquiring, a moving target: the emerging models for the governance of (personal big) data.

The review covered documents, publications, news and websites in English that addressed emerging practices for the governance of data with a focus on the European context. On the whole, it included 72 academic articles, 16 book chapters, 63 reports and policy documents, and 22 websites of projects/initiatives. The resources were collected in the time span from October 2018 to July 2019, with nine documents added during the review process. Most of these are recent, as 74% have been published from 2017 onwards. Scholarship on data from a ‘Global South’ perspective (Arora, 2019; Milan and Treré, 2019) suggests that different geographical, political, social, organisational, and jurisdictional contexts also affect roles and power in the (data) governance discourse. This contribution however, primarily takes a European standpoint in investigating data governance models. In doing so, we acknowledge that we may have missed insights deriving from other geographical arenas and that our models need to be read as relative to the European context.

Analytical dimensions

To guide our analysis and description of the emerging models of data governance we used the following analytical dimensions, drawing in particular from Abraham et al. (2019) and Winter and Davidson (2019) (see Table 1). These dimensions relate with STS and CDS by bringing to the fore the interests and goals of the main stakeholders involved in data governance, and – helping us to make visible the power relations and the different forms of agency in each model.

Table 1. Analytical dimensions.

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

Stakeholders

Stakeholders are all actors, such as individuals, organisations and groups, who are affected by, or have an effect on, the way data is governed and the value that is created from it.. Stakeholders differ widely in terms of possibilities to access, control and process data, as well as knowledge about how data is collected and treated. They also hold different values and interests about data, and norms regarding its use (Winter and Davidson, 2019). They include ‘data subjects, data controllers (and processors), and third-party data users’ (Ho and Chuangt, 2019: 203). Stakeholders encompass private sector, public sector, academia, scientific and civic organisations, activists, social entrepreneurs and citizens (Calzada, 2017).

Governance goals

Governance goals are value-based objectives that different stakeholders have established for governing data (Winter and Davidson, 2019). These goals are the meanings data represents for the interested actors. While some goals may be broadly acknowledged or even shared within and across use contexts, others might be opaque or disputed between different actors or contexts. A straightforward goal for companies is to maximise financial returns through data sharing and aggregation (Srnicsek, 2017). Policy documents, instead, might cite public interest as one of the key goals to pursue with data sharing agreements. Another goal for data governance could be increasing data subjects’ control of their data or giving voice to disadvantaged groups (Beraldo and Milan, 2019; Winter and Davidson, 2019).

Value from the data

It refers to the kind of value that is created from data through aggregation, analytics, and business intelligence, by the various stakeholders who may reap different benefits from these processes (Winter and Davidson, 2019). The value that stakeholders gain

varies, from economic revenues, to public good and citizens’ self-determination. This dimension assesses to what extent a model foresees that data is used in ‘socially progressive ways’ (Kitchin and Laurialt, 2014). Is any form of public value created or only private value for companies and/or the individual users? Through this dimension we enquire whether the value and knowledge produced through data aggregation and analysis is redistributed between actors and across society (Mulgan and Straub, 2019).

Governance mechanisms

With governance mechanisms we refer to the strategies and instruments adopted by different agents to achieve their goals and direct change in a socio-technical system (Borrás and Edler, 2014). This dimension comprises the elements of a data assemblage (Kitchin, 2014) that frame how data is controlled, what value is created, and who benefits from it. It includes: system of thoughts, policies, regulations, committees, contracts, terms of service, standards, algorithms, interfaces and other socio-technical systems that form the governance mechanisms of today’s data infrastructures, and that some actors can exploit better than others (Kitchin, 2014; Abraham et al., 2019; Winter and Davidson, 2019). Informed by STS, the notion of governance mechanisms include also the broader ethical, political and economic principles embedded into the data infrastructures and represent the complex socio-technical ‘assemblage’ in which data governance takes place (Bowker et al., 2010; Slota and Bowker, 2017).

Reciprocity

Reciprocity refers to the power relations between stakeholders in accessing, controlling and using data. It highlights the difference between unilateral approaches, such as those in which big tech corporations hold most of the decision-making power, to mutual data governance models in which more stakeholders take part in the governance of data. This

Table 2. Summary of data governance models.

Model	Key actors	Goals	Value	Mechanisms
Data sharing pools (DSPs)	<ul style="list-style-type: none"> • Business entities • Public bodies 	<ul style="list-style-type: none"> • Fill knowledge gaps through data sharing • Innovate and develop new services 	<ul style="list-style-type: none"> • Private profit • Economic growth 	<ul style="list-style-type: none"> • Principle of 'data as a commodity' • Partnerships • Contracts (e.g. repeatable frameworks)
Data cooperatives (DCs)	<ul style="list-style-type: none"> • Civic organisations • Data subjects 	<ul style="list-style-type: none"> • Rebalance power unbalances of the current data economy • Address societal challenges • Foster social justice and fairer conditions for value production • Inform policy-making • Address societal challenges • Innovate • Adopt a responsible approach to data 	<ul style="list-style-type: none"> • Public interest • Scientific research • Empowered data subjects 	<ul style="list-style-type: none"> • Principles from the cooperative movement • Data commons • 'Bottom-up' data trusts • GDPR Right to data portability
Public data trusts (PDTs)	<ul style="list-style-type: none"> • Public bodies 	<ul style="list-style-type: none"> • Inform policy-making • Address societal challenges • Innovate • Adopt a responsible approach to data 	<ul style="list-style-type: none"> • Public interest • More efficient public service delivery 	<ul style="list-style-type: none"> • Principle of 'data as a public infrastructure' • Trust building initiatives • Trusted intermediaries • Enabling legal framework
Personal data sovereignty (PDS)	<ul style="list-style-type: none"> • Business entities • Data subjects 	<ul style="list-style-type: none"> • Data subjects self-determination • Rebalance power unbalances of the current data economy • Develop new digital services • centred on users need 	<ul style="list-style-type: none"> • Empowered data subjects • Economic growth • Private profit • Knowledge 	<ul style="list-style-type: none"> • Principle of 'technological sovereignty' • Communities and movements (e.g. MyData) • Intermediary digital services (personal data spaces) • GDPR Right to data portability

dimension links with the notion of data politics that understands data as ‘generative of new forms of power relations’ (Ruppert et al., 2017: 2).

Emerging data governance models

This section describes the data governance models identified following the five dimensions described above. These models should be understood as ideal types in the Weberian sense. They are analytical constructs that emphasise certain traits in order to synthesise phenomena that differ for the degree of affiliation to those traits (Kvist, 2007). They are not intended as an exhaustive description of the state of the art, but as a contribution in synthesising emerging data governance models. The analysis includes models that differ, to varying degrees, from the current dominant one. Therefore, we do *not* account for cases in which platforms engage in data sharing with other actors, but retain full control over data, deciding unilaterally which other stakeholders to bring inside, what data they can access and what they can do with it (Shkabatur, 2019). The four models described are labelled: data sharing pools (DSPs), data cooperatives (DCs), public data trusts (PDTs) and personal data sovereignty (PDS).

Data sharing pools

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

Governance mechanisms for DSPs include technical architectures, such as data sharing platforms and Application Programming Interfaces (APIs), which facilitate a centralised data exchange within business ecosystems. However, a key mechanism is the contract,

a legal and policy framework, that defines the modalities for data sharing, how data can be handled, and for which purposes. These contracts could be ‘repeatable frameworks of terms and mechanisms to facilitate the sharing of data’ between entities, which are especially useful for organisations that do not have the know-how and legal support to leverage data (Hall and Pesenti, 2017; Hardings and Wells, 2018). Although these frameworks have been referred to as data trusts, there is not a full consensus whether they could be assimilated to actual legal trust structures or a ‘marketing tool’ facilitating the responsible sharing of data (Delacroix and Lawrence, 2019: 242).

An example of DSP is the Connected Citizens Program, a collaboration between Waze, a community-based traffic and transport app, E.S.R.I., a global commercial software company, and municipal governments (Shkabatur, 2019). As part of the pool, municipal governments share real-time construction and road closure data through the E.S.R.I. platform, and in exchange Waze² shares its community-collected real-time traffic data.³ The assumption of this kind of contracts is that all parties benefit since the DSP enables them to obtain easily data that would otherwise be inaccessible. There is reciprocity between partner organisations, but only data holders are involved, as data subjects tend to be excluded from the relation and are at best depicted as passively benefiting from it. Although use cases of DSPs do exist, examples in practice are still few (Mattioli, 2017). A practical limitation consists in the transaction costs, such as data preparation, ensuring privacy and interoperability challenges, which put small businesses and underfunded entities at a disadvantage (GovLab, 2018). A further limitation is that often there is one dominant partner (Carballa Smichowski, 2019). Therefore, although involving potentially many actors beyond big tech platforms, the relations are not necessary as horizontals (and sustainable) as claimed.

Data cooperatives

DCs distribute data access/rights among actors like DSPs, but differently from those, provide higher involvement of data subjects and are guided by different goals. DCs enable a de-centralised data governance approach in which data subjects ‘voluntarily pool their data together, to create a common pool for mutual benefits’ (Ho and Chuangt, 2019: 204). Participants of DCs share data while retaining control over it, having a say on how it is managed and put to value, and not submitting to the extractive logic of digital capitalism (Borkin, 2019; Ho and Chuangt, 2019). Therefore, data subjects are key stakeholders within DCs. By establishing a relationship of trust with the

cooperative that manages data on their behalf, they preserve democratic control over their data and might demand an equitable share in the benefits produced (Borkin, 2019; Delacroix and Lawrence, 2019). This model is characterised by high reciprocity since ‘all parties are stakeholders and are equally affected and bound by the governing rules they discuss, negotiate and then agree upon’ (Ho and Chuangt, 2019: 203).

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

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

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

Public data trusts

PDTs refer to a model of data governance in which a public actor accesses, aggregates and uses data about its citizens, including data held by commercial entities, with which it establishes a relationship of trust (Delacroix and Lawrence, 2019; Hall and Pesenti,

2017; Mulgan and Straub, 2019). Several stakeholders might be involved in this model, including city administrators, managers of public institutions, platform companies, trusted data intermediaries, research institutions, start-ups, and SMEs. Public administrations may also invite third-parties to access their data sources and develop data-driven services and/or to offer guidance on data sharing (Hall and Pesenti, 2017; Morozov and Bria, 2018). A key goal of PDTs is to integrate data from multiple sources to inform policy-making, promote innovation and address societal challenges, while adopting a responsible approach to the use of personal data (Bass et al., 2018; Collinge, 2016; Kawalek and Bayat, 2017; Morozov and Bria, 2018; Van Zoonen, 2016).

In PDTs, public actors assume the role of trustees that guarantee citizens’ data is handled ethically, privately and securely. Thus they imply the establishment of a relationship of trust between citizens and public bodies: citizens must be reassured that public actors are capable to keep their personal information safe and secure and that they will use data to improve their lives (Collinge, 2018). To earn such level of trust from citizens, public bodies might engage in citizens’ consultations and living labs, or require the intervention of external independent organisations that act as trusted intermediaries (Collinge, 2018; EC, 2020c; Mulgan and Straub, 2019). These trusted intermediaries are new institutions that are allegedly held to account for securely managing data, preserving citizens’ privacy, and maximising the public value of data (Mulgan and Straub, 2019). These entities will be independent and unrelated to for-profit firms and big tech corporations, and guarantee that data is managed without abuses through strong accountability and standards.⁴ Therefore, even if citizens are mostly seen as recipients who benefit from services and policies developed through PDTs, they might be explicitly involved in this model through ‘trust building’ governance mechanisms such as living labs, public consultations and civic society initiatives.

Examples of PDTs are pilot projects by the Open Data Institute, a non-profit private company, in conjunction with the Mayor of London and the Royal Borough of Greenwich. These projects use real-time data to improve public service delivery, such as in council-owned social houses and public parking (Open Data Institute, 2019). The city of Barcelona has also been experimenting with PDTs including ‘clauses within procurement contracts specifying that a service provider must make any data that may be of public value available to the city council’ (Bass et al., 2018: 28).

An underlying assumption of PDTs is that all data with a public interest component (even if collected by

commercial entities) is part of a nation infrastructure (National Infrastructure Commission, 2017) therefore the information it affords should be ‘socialised’ to produce value for citizens and society as a whole (Cardullo, 2019; De Lange, 2019; Morozov and Bria, 2018). Currently, the involvement of private companies in such forms of data sharing takes place only on a voluntary basis while government-owned and utility companies (such as energy and transport) have more motives to collaborate with public bodies (Bass et al., 2018; Open Data Institute, 2019).

Whilst at present PDTs are largely limited to small pilot projects, a key enabler would be a legal framework mandating private companies to grant access to data of public interest to public actors under conditions specified in the law (Shkabatur, 2019). This was considered by the EC (2020c), which then appointed a High-Level Expert Group on Business-to-Government data sharing. The issue has also been discussed at national level in Europe. For instance, French Member of Parliament Belot proposed creating the legal concept of ‘territorial interest data’ to give local governments the power to demand access to data (Carballa Smichowski, 2019).

Personal data sovereignty

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

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

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

PDS has been especially encouraged within the context of MyData, an international movement and a community of activists, non-profit organisations, think-tanks as well as commercial actors, start-ups and SMEs. An analysis of this movement (Lehtiniemi and Haapoja, 2020) highlights its inherent tensions between activists’ interests for social change and the economic interests of commercial firms. The same tension stands at the core of the PDS model and its positioning towards value generation. PDSs are expected to produce value in the form of data subjects’ self-determination, knowledge, and public interest, but at the same time foster economic growth through an ecosystem of new commercial services supporting them.

A limit of this model lies in its dependence on personal data spaces as these are currently adopted by only a niche of users and often fail to scale beyond pilots (Ilves and Osimo, 2019). As, as business entities, they may have interested in how to ‘nudge’ users and a few personal data spaces might gain more power in the market (Lehtiniemi, 2017). Furthermore, citizens have limited awareness about platforms’ use of personal data for profit and the need for alternative models of value production, and the majority would not be capable, nor have the time to, take advantage of the opportunities offered by these intermediary services (e.g. Andrejevic, 2014). Envisioning citizens as ‘market agents’ (Lehtiniemi and Haapoja, 2020) free to choose from an ecosystem of personal data spaces might not fully address the asymmetries of power of the current data landscape.

Discussion

In this article we contribute to the literature and the policy debate on data governance using a socio-technical perspective to describe four emerging models of data governance: DSPs, DCs, PDTs and PDS. The models are abstract conceptualisations (Kvist, 2007) that do not necessarily represent discrete implementations of data governance. A single initiative could embrace more than one of these conceptual

models simultaneously, or be inspired by one but not embrace it fully, as reality is messier than abstract constructs. Nonetheless, they provide a foundation for discussion on alternative approaches or ‘desirable futures’ for accessing and sharing data in the age of datafication (Jasanoff, 2015). They could be read as inventive practices that problematise current arrangements and reassemble them in accordance to the interests of the actors involved. Table 2 presents a summary of the main features of the models on the five analytical dimensions that guided our analysis.

All models highlight a concern for redressing the structural power imbalances between corporate platforms and other actors, such as data subjects, public bodies, third parties, civil society and researchers. There are nonetheless substantial differences regarding which stakeholders exert influence over data, and what value is pursued through data use. Drawing from the notions of data infrastructure and data politics, we highlighted the plurality of actors that affects or is affected by the way data is made accessible and used in each model. The actors’ roles and their power to control data are situated and contingent: they relate to the broader ethical, legal, political and economic principles that are embedded in the data infrastructure and the various governance mechanisms that enable each model. The governance goals of more powerful actors both support and are supported by the ‘imaginaries’ that prevail in each model, which in turn influence value generation and redistribution.

In DSPs, one of the classic rhetoric of Big Data is embraced: data creates more value if aggregated. In that spirit, two or more data holders (both private and public) join forces and establish data sharing agreements. They analyse each other’s data filling knowledge gaps and fostering data-driven innovation. On the surface this model promotes reciprocity between, potentially many, data holders, as it is based on horizontal relationships. Yet, it also fosters power asymmetries. Data holders with more resources or that possess more valuable datasets have greater power to set the terms on how data is accessed and used. Furthermore, data subjects (and citizens in general) do not have a voice in this model; they are not included in the relation and are at best depicted as recipients of the innovations developed through it.

PDS on the contrary place data subjects at the centre of an ecosystem of new services that provide them the means to access, control, share and analyse their data. Based on the principle of sovereignty, this model emphasises individual control over data and self-determination, and it is in stark contrast to surveillance capitalism (Lehtiniemi, 2017). A movement of data activists is promoting this progressive goal, at the same time commercial actors are interested in it

as a means to support an ecosystem of new services. In PDS different kinds of actors, with different interests, converge for the promotion of a ‘fairer data economy’ (Ilves and Osimo, 2019). This tension leads to some inconsistencies. A fully private ecosystem of intermediary services, even if allowing users to control their data, would leave the incumbency of addressing the power asymmetries of the current data landscape to the market (Lehtiniemi and Haapoja, 2020). From studies on citizens’ perspectives on their digital data, it also appears that the majority would not have the skills and interest to take advantage of the opportunities offered by these intermediary services (e.g. Andrejevic, 2014).

A wider range of actors is involved in PDTs, with public bodies taking the lead. The underlying principle of PDTs is that all data with a public interest component is part of a nation infrastructure, and therefore the information it affords should be ‘socialised’ to produce value for citizens and society as a whole. Actors from public sector, non-profit, business and academia take part in PDTs. Data subjects, however, are not just recipients of the services developed through data. To be effective, PDTs imply the establishment of a relationship of trust with citizens who must be reassured their personal information is protected and that will be used for the public interest. To achieve this objective, public bodies need to build trust listening to requests from civic society. As recent failed initiatives (Sidewalk Toronto), as well as the debate for COVID-19 contact tracing apps (Ada Lovelace Institute, 2020), show this is something that public bodies are (still) learning to do. Another challenge for PDTs is to establish data sharing agreements with private companies, which might have data of public interest but not be willing to share it unless regulations mandate it.

DCs are a grassroots-driven decentralised data governance model in which citizens voluntarily pool their data together establishing a relationship of trust with the cooperative that manages data on their behalf. Data subjects preserve democratic control over data and have an equitable share in the benefits produced, which are often aimed at the public interest (such as medical research). Drawing from the principles of the cooperative movement and platform cooperatives, this model is in stark opposition to platform capitalism and aims to be a fairer, transparent and non-monopolistic alternative. Big tech platform might be completely excluded from data governance in a DC – even if informally present as the ‘antagonistic actor’ – or be included only as a data provider from which users take their data. Although DC raise a lot of interest in the policy debate as an ethical approach to data sharing, they struggle to scale up and reach a critical mass of users (Sandoval, 2020).

With respect to the kind of value pursued, we see that DPS mainly focus on producing economic value, while other forms of value gradually ‘chime in’ in the remaining models, such as social change, public interest, fairness, and data subjects’ self-determination. It should also be noted that for the most part, these models could be found in niche initiatives or pilot projects, and there is still limited research concerning the value they generate and their sustainability over time (Borkin, 2019; Verhulst, 2019). Value production and redistribution, thus, can be assessed more at the level of the imaginary, than from evaluations of tangible outcomes. Nevertheless, adopting a normative perspective informed by CDS, we could ask to what extent these data governance models foster a redistribution of value.

In DSPs, data is a ‘market commodity’ and economic value is redistributed horizontally among data holders who join the partnership. PDS put forward important innovations for data subjects’ exerting digital rights, but do not question the datafication and commodification mechanisms of the platform society (Van Dijck et al., 2018). They are oriented towards the creation of value for the individual (self-determination) and new commercial actors (data services), with public interest as a by-product of these. The remaining models expressively pursue the public interest: DCs allow data subjects to collect and aggregate their data for the public interest, while PDTs act on behalf of citizens, aggregating and analysing different data sources to inform policy-making and address societal challenges. If in DCs a cooperative has to be trusted, in PDTs is a public body. Yet, in the latter, it might also be that a trusted external independent organisation acts as a data intermediary between citizens and a public body; this demonstrates how the abstract models can easily overlap in practice.

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

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

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

A final consideration concerns the intertwined relationship between the data practices we have examined and the regulatory frameworks in which they exist. These data governance models can only develop further if they are sustained by appropriate legal frameworks, such as the GDPR for personal data or a new legal act to mandate access to commercial data of public interest. With the recent developments in data policy (EC, 2020a), the EC is strengthening its role as transnational regulator of technology with repercussions on a global scale. In doing so, it will be crucial

to engage with the wider set of stakeholders identified in our research including local administrations and many actors from civil society who have an important role in shaping the emerging forms of data governance that address the asymmetries of the current data landscape.

Limitations

The lack of established terms in the field, as well as the ever-changing nature of the theme, has complicated the selection of documents. The procedure adopted for retrieving resources, detailed in the methodology section, is not entirely systematic, hence the findings have to be contextualised in that approach, including the time span in which the research has been conducted and the European focus. Furthermore, the described models do not claim to represent the full spectrum of emerging models currently developed for the governance of data. Being ideal types, they have a heuristic value as tools to be adopted for further studies.

Another limitation relates to our multi-dimensional framework that derives from our research questions. In focusing exclusively on the dimensions of stakeholders, governance goals, value from the data, governance mechanisms, and reciprocity, we acknowledge that many more lenses could have been adopted, such as, for example, the perspectives of trust and distrust, privacy and data protection, authority and authorities, law and regulations. We selected the dimensions we thought were the most appropriate to examine emerging data practices from a social science perspective, yet we acknowledge other important elements might have been included.

A final limitation is that we do not include primary data collected to investigate what lays behind the *façade* of practitioners' literature. We are indeed addressing this aspect through empirical research as a means to examine the characteristics of some of these models and the challenges for their implementation, largely in an urban context. This study in fact aims to lay the foundation for future investigations of how socio-technical assemblages unfold around different manifestations of data governance. How far these models work in practice? To what extent are these replicable in different contexts? What are the key enablers and obstacles? This type of microanalysis could help recognise which mechanisms are used (or lacking) in each particular context and which strategic practices are essential.

Conclusions

Notwithstanding its limitations, this study shows that many actors – from public sector, academia,

businesses, civic society, as well as activists and social entrepreneurs – are seeking alternatives to the dominant data governance model. We discussed four of the models that emerge from the practices of these actors. The social practices for data access, sharing, control and use, and the derived models come at a crucial time as the discussion on data governance and data sovereignty is building a new momentum in Europe with the emergence of AI as a strategic area of policy for the future of society. Therefore, we believe it is (and will be) important to examine data governance arrangements in a rapid and timely manner (Mann et al., 2019), and understand how to incentivise uses of data at the service of the public good. To do so, it might be helpful to adopt also a social science perspective on data governance that allows seeing 'through the infrastructure' to ask: what principles guide data sharing and use? What is done with data and who can access and participate in its governance? What value is produced and how it is redistributed? Ultimately, this article wishes to encourage more 'normative conversations' on socio-technical imaginaries and 'desirable futures': the kind of society we want to live in and how we can shape the digital transformation accordingly (Jasanoff, 2015; Kitchin, 2014).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Acknowledgements

We thank the anonymous reviewers and the editors for their thoughtful constructive feedback that greatly improved the article. We also thank Malte Beyer, Jutta Thielen Del Pozo, Bertin Martens and Andrea Perego of the European Commission who gave feedback on an earlier version of the manuscript. The views expressed in this article are solely the responsibility of the authors and do not reflect the official opinion of the European Commission.

ORCID iD

Marina Micheli  <https://orcid.org/0000-0002-7004-6480>

Notes

1. We searched the Web of Science using the term 'data governance' from 2014 through June 2019 and the categories: management OR communication OR social sciences interdisciplinary OR law OR sociology. The 26 documents

resulted from this search, however, were excluded from further consideration because not pertinent to the focused purpose of this review. Results from Google Scholars for 'data governance' are numerous, but mainly produced within other disciplines and fields, such as library information, organisational studies, management, and business.

2. <https://www.waze.com/ccp>
3. <https://datacollaboratives.org/cases/esri-and-waze-open-data-sharing-for-governments.html>
4. Differently from the case of Toronto Quayside project that sparked many controversies because the civic data trust was commissioned to Sidewalk Labs, a subsidiary of Google parent company Alphabet, hence privatising public data.

References

- Abraham R, Schneider J and Vom Brocke J (2019) Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management* 49: 424–438.
- Ada Lovelace Institute (2020) *Exit through the App Store?* Rapid evidence review, Ada Lovelace Institute, London, UK, April.
- Andrejevic M (2014) The big data divide. *International Journal of Communication* 8: 1673–1689.
- Arora P (2019) Decolonizing privacy studies. *Television & New Media* 20(4): 366–378.
- Bass T, Sutherland E and Symons T (2018) *Reclaiming the Smart City. Personal data, trust and the new commons*. Report, NESTA, London, UK, July.
- Beer D (2017) The social power of algorithms. *Information, Communication & Society* 20(1): 1–13.
- Bennett CJ and Raab CD (2018) Revisiting the governance of privacy: Contemporary policy instruments in global perspective. *Regulation & Governance* 1: 18.
- Beraldo D and Milan S (2019) From data politics to the contentious politics of data. *Big Data & Society* 6(2): 1–11.
- Bigo D, Isin E and Ruppert E (2019) *Data Politics: Worlds, Subjects, Rights*. New York, NY: Routledge.
- Borkin S (2019) *Platform co-operatives. Solving the capital conundrum*. Report, NESTA and Co-operatives UK, London, UK, February.
- Borrás S and Edler J (2014) The Governance of change in socio-technical and innovation systems: Three pillars for a conceptual framework. In: Borrás S and Edler J (eds) *The Governance of Socio-Technical Systems: Explaining Change*. Cheltenham: Edward Elgar Publishing, pp. 23–48.
- Bowker GC, Baker K, Millerand F, et al. (2010) Toward information infrastructure studies: Ways of knowing in a networked environment. In: Hunsinger J, Klastrop L and Allen M (eds) *International Handbook of Internet Research*. Dordrecht: Springer, pp. 97–117.
- Calzada I (2017) From smart cities to experimental cities? In: Giorgino VM and Walsh ZD (eds) *Co-Designing Economies in Transition: Radical Approaches in Dialogue with Contemplative Social Sciences*. London: Palgrave Macmillan, pp. 191–217.
- Carballa Smichowski B (2019) Alternative data governance models: Moving beyond One-Size-Fits-All solutions. *Intereconomics. Review of European Economic Policy* 54(4): 222–227.
- Cardullo P (2019) Smart commons or a “Smart Approach” to the commons? In: Cardullo P, Di Felicianantonio C and Kitchin R. (eds) *The Right to the Smart City*. Bingley: Emerald Publishing Limited, pp. 85–98.
- Colebatch HK (2014) Making sense of governance. *Policy and Society* 33(4): 307–316.
- Collinge A (2016) Data for London – A city data strategy for a smart city future. In: London Data Store. Available at: <https://data.london.gov.uk/blog/data-for-london-a-city-data-strategy-for-a-smart-city-future/> (accessed 18 June 2020).
- Collinge A (2018) A ‘new deal’ for city data? In: London Data Store. Available at: <https://data.london.gov.uk/blog/a-new-deal-for-city-data/> (accessed 18 June 2020).
- Couldry N and Powell A (2014) Big data from the bottom up. *Big Data & Society* 1(2): 1–5.
- Couture S and Toupin S (2018) What does the concept of ‘sovereignty’ mean in digital, network and technological sovereignty? In: *GigaNet: Global internet governance academic network, annual symposium 2017*. Available at: <https://ssrn.com/abstract=3107272> (accessed 18 June 2020).
- Craglia M, Annoni A, Benczur P, et al. (2018) *Artificial Intelligence: A European Perspective*. Report, Joint Research Centre. Publications Office, Luxembourg.
- De Lange M (2019) The right to the datafied city: Interfacing the urban data commons. In: Cardullo P, Di Felicianantonio C and Kitchin R (eds) *The Right to the Smart City*. Bingley: Emerald Publishing Limited, pp. 71–83.
- Delacroix S and Lawrence ND (2019) Bottom-up data trusts: Disturbing the ‘one size fits all’ approach to data governance. *International Data Privacy Law* 9(4): 236–252.
- DeNardis L (2019) The social media challenge to internet governance. In: Graham M and Dutton WH (eds) *Society and the Internet: How Networks of Information and Communication are Changing Our Lives*. Oxford: Oxford University Press, pp. 389–402.
- European Commission (EC) (2018a) *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions “Artificial Intelligence for Europe” COM/2018/237 Final*. Publications Office, Luxembourg.
- European Commission (EC) (2018b) *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions “Coordinated Plan on Artificial Intelligence” COM/2018/795 Final*. Publications Office, Luxembourg.
- European Commission (EC) (2020a) *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions “Shaping Europe’s Digital Future” COM(2020) 67 Final*. Publications Office, Luxembourg.

- European Commission (EC) (2020b) *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions "a European Strategy for Data"* COM(2020) 66 Final. Publications Office, Luxembourg.
- European Commission (EC) (2020c) *Towards a European Strategy on Business-to-Government Data Sharing for the Public Interest. Final Report Prepared by the High-Level Expert Group on Business-to-Government Data Sharing*. Publication Office, Luxembourg.
- Gorwa R and Garton Ash T (2020). Democratic Transparency in the Platform Society. In: Persily N, Tucker J (eds) *Social Media and Democracy: The State of the Field and Prospects for Reform*. Cambridge, UK: Cambridge University Press, pp. 286–12.
- GovLab (2018) Data Stewards: Data leadership to address 21st century challenges. In: Medium. Available at: <https://medium.com/@TheGovLab/data-stewards-data-leadership-to-address-21st-century-challenges-a5e555cd4235> (accessed 18 June 2020).
- Gray J, Gerlitz C and Bounegru L (2018) Data infrastructure literacy. *Big Data & Society* 5(2): 1–13.
- Hall W and Pesenti J (2017) *Growing the artificial intelligence industry in the UK*. Independent report, UK.
- Hardingens J and Wells P (2018) Defining a 'data trust'. In: Open Data Institute Blog. Available at: <https://theodi.org/article/defining-a-data-trust/> (accessed 18 June 2020).
- Harrison S and Mort M (1998) Which champions, which people? Public and user involvement in health care as a technology of legitimization. *Social Policy & Administration* 32(1): 60–70.
- Heeks R and Shekhar S (2019) Datafication, development and marginalised urban communities: An applied data justice framework. *Information, Communication & Society* 22(7): 992–1011.
- Ho C and Chuang T (2019) Governance of communal data sharing. In: Daly A, Devitt K and Mann M (eds) *Good Data*. Amsterdam: Institute of Network Cultures, pp. 202–213.
- Ilves LK and Osimo D (2019) *A roadmap for a fair data economy. Policy Brief, Sitra and the Lisbon Council*. Helsinki, Finland, April.
- Jasanoff S (2015) Future imperfect: Science, technology, and the imaginations of modernity. In: Jasanoff S and Kim SH (eds) *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. Chicago, IL: University of Chicago Press, pp. 1–33.
- Kawalek P and Bayat A (2017) *Data as Infrastructure*. National Infrastructure Commission. Available at: <https://www.nic.org.uk/publications/data-infrastructure-paper-national-infrastructure-commission> (accessed 18 June 2020).
- Kennedy H and Moss G (2015) Known or knowing publics? Social media data mining and the question of public agency. *Big Data & Society* 2(2): 1–11.
- Kennedy H, Poell T and Van Dijck J (2015) Introduction: Special issue on data and agency. *Big Data & Society* 2(2): 1–7.
- Khatri V and Brown CV (2010) Designing data governance. *Communications of the ACM* 53(1): 148–152.
- Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.
- Kitchin R (2017) Thinking critically about and researching algorithms. *Information, Communication & Society* 20(1): 14–29.
- Kitchin R and Laurialt T (2014) Towards critical data studies: Charting and unpacking data assemblages and their work. In: Thatcher J, Eckert J and Shears A (eds) *Thinking Big Data in Geography. New Regimes, New Research*. Lincoln/London: University of Nebraska Press, pp. 83–94.
- Kooiman J (2003) *Governing as Governance*. London: Sage.
- Kvist J (2007) Fuzzy set ideal type analysis. *Journal of Business Research* 60(5): 474–481.
- Lehtiniemi T (2017) Personal data spaces: An intervention in surveillance capitalism?. *Surveillance & Society* 15(5): 626–639.
- Lehtiniemi T and Haapoja J (2020) Data agency at stake: MyData activism and alternative frames of equal participation. *New Media & Society* 22(1): 87–104.
- Mann M and Devitt SK and Daly A (2019) What Is (in) Good Data? In: Daly A, Devitt K and Mann M (eds) *Good Data*. Amsterdam: Institute of Network Cultures, pp. 8–23.
- Mattioli M (2017) The data-pooling problem. *Berkeley Technology Law Journal* 32: 179–236.
- Micheli M, Blakemore M, Ponti M, et al. (2018) *The governance of data in a digitally transformed European Society*. Second Workshop of the DigiTranScope Project, European Commission, JRC114711.
- Milan S and Treré E (2019) Big data from the South(s): Beyond data universalism. *Television & New Media* 20(4): 319–335.
- Morozov E and Bria F (2018) *Rethinking the Smart City: Democratizing Urban Technology*. New York, NY: Rosa Luxemburg Stiftung.
- Mulgan G and Straub V (2019) *The new ecosystem of trust: How data trusts, collaboratives and coops can help govern data for the maximum public benefit*. Nesta. Available at: <https://www.nesta.org.uk/blog/new-ecosystemtrust/> (accessed 18 June 2020).
- National Infrastructure Commission (2017) *Data for the public good*. Report, National Infrastructure Commission. Available at: <https://www.nic.org.uk/wp-content/uploads/Data-for-the-Public-Good-NIC-Report.pdf> (accessed 18 June 2020).
- Open Data Institute (2019) *Data trusts: Lessons from three pilots*. Report, Open Data Institute, London, UK. Available at: <https://theodi.org/article/odi-data-trusts-report/> (accessed 20 June 2020).
- Pazaitis A, Kostakis V and Bauwens M (2017) Digital economy and the rise of open cooperativism: The case of the enspiral network. *Transfer: European Review of Labour and Research* 23(2): 177–192.
- Rhodes RAW (1996) The new governance: Governing without government. *Political Studies* 44(4): 652–667.
- Rosenbaum S (2010) Data governance and stewardship: Designing data stewardship entities and

- advancing data access. *Health Services Research* 45(5 Pt2): 1442–1455.
- Ruppert E, Isin E and Bigo D (2017) *Big Data & Society* 4(2): 1–7.
- Sandoval M (2020) Entrepreneurial activism? Platform cooperativism between subversion and co-optation. *Critical Sociology* 46(6): 801–817.
- Scholz T (2016) *Platform cooperativism. Challenging the corporate sharing economy*. Report for the Rosa Luxemburg Foundation, New York, USA.
- Shkabatur J (2019) The global commons of data. *Stanford Technology Law Review* 22: 1–46.
- Slota SC and Bowker GC (2017) How infrastructure matters. In: Felt U, Fouche R, Miller CA and Smith-Doerr L (eds) *The Handbook of Science and Technology Studies (4th edition)*. London: MIT Press, pp. 529–554.
- Srnicek N (2017) *Platform Capitalism*. Cambridge: Polity Press.
- Symons T and Bass T (2017) *Me, my Data and I: The Future of the Personal Data Economy*. Report, NESTA, London, UK, September.
- Taylor L (2017) What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society* 4(2): 1–14.
- Taylor L and Broeders D (2015) In the name of development: Power, profit and the datafication of the global South. *Geoforum* 64: 229–237.
- Van Asselt M and Renn O (2011) Risk governance. *Journal of Risk Research* 14(4): 431–449.
- Van Dijck J, Poell T and De Waal M (2018) *The Platform Society: Public Values in a Connective World*. Oxford: Oxford University Press.
- Van Zoonen L (2016) Privacy concerns in smart cities. *Government Information Quarterly* 33(3): 472–480.
- Verhulst SG (2019) Data Collaboratives as an enabling infrastructure for AI for Good. In: Qatar Computing Research Institute. Available at: <https://qcai.qcri.org/index.php/2019/04/15/data-collaboratives-as-an-enabling-infrastructure-for-ai-for-good/> (accessed 18 June 2020).
- Villani C (2018) For a meaningful artificial intelligence: Towards a French and European strategy, Mission assigned by the Prime Minister Édouard Philippe.
- Winter JS and Davidson E (2019) Big data governance of personal health information and challenges to contextual integrity. *The Information Society* 35(1): 36–51.
- Wolf KD (2002) Contextualizing normative standards for legitimate governance beyond the state. In Grote JR and Gbikpi B (eds) *Participatory Governance: Political and Societal Implications*. Opladen: Leske und Budrich, pp. 35–50.
- Zuboff S (2015) Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30(1): 75–89.