

The value of Big Data in government: The case of ‘smart cities’

Big Data & Society
January–June 2020: 1–14
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2053951720912775
journals.sagepub.com/home/bds



Karl Löfgren¹ and C William R Webster²

Abstract

The emergence of Big Data has added a new aspect to conceptualizing the use of digital technologies in the delivery of public services and for realizing digital governance. This article explores, via the ‘value-chain’ approach, the evolution of digital governance research, and aligns it with current developments associated with data analytics, often referred to as ‘Big Data’. In many ways, the current discourse around Big Data reiterates and repeats established commentaries within the eGovernment research community. This body of knowledge provides an opportunity to reflect on the ‘promise’ of Big Data, both in relation to service delivery and policy formulation. This includes, issues associated with the quality and reliability of data, from mixing public and private sector data, issues associated with the ownership of raw and manipulated data, and ethical issues concerning surveillance and privacy. These insights and the issues raised help assess the value of Big Data in government and smart city environments.

Keywords

Big Data, smart cities, public value chain, value chain analysis, eGovernment, privacy

This article is a part of special theme on Big Data and Surveillance. To see a full list of all articles in this special theme, please click here: <https://journals.sagepub.com/page/bds/collections/hypecommerciallogics>

Introduction

It is widely reported that data analytics, or ‘Big Data’, is going to radically transform society. Whilst there are multiple definitions of Big Data (Manzoor, 2015; Ylijoki and Porras, 2016) the vision of new digitally oriented Big Data practices to collect, mine, store and process data in radical new ways has become well established and is closely associated with visions about the subsequent transformation of public policy and service delivery. Practices associated with Big Data, especially around machine learning, automated decision-making and predictive algorithms, are changing how public service decision-makers and providers envisage future technology solutions in all service arenas (Mayer-Schonberger and Cukier, 2013), including those in smart city urban environments. Without neglecting the implementation of data analytics in governmental settings and the realization of genuine benefits (for some empirical examples see: Gamage, 2016), the current discourse reflects a technocratic agenda based on a scientific-technological rationality

(Esmark, 2016). Typically, urban public service delivery is aligned with commercial ambitions to harvest data from citizens and public values such as transparency and fairness are outranked by attention to instrumental values such as efficiency, safety and security (Meijer et al., 2016). In terms of governance, there is an alignment with a corporate discourse where the private sector represents the ‘gold-standard’ of technological deployment to be replicated in the public sector. This approach neglects the integrity of the public sector, its institutional norms and values, and the uniqueness of the public organization in terms of safeguarding the rule

¹School of Government, Victoria University of Wellington, New Zealand

²Centre for Research into Information, Surveillance and Privacy (CRISP), Stirling Management School, University of Stirling, Scotland, UK

Corresponding author:

C William R Webster, Centre for Research into Information, Surveillance and Privacy (CRISP), Stirling Management School, University of Stirling, UK.

Email: william.webster@stir.ac.uk



of law, political neutrality, democratic control, accountability, and the assurance of other non-economic public values (cf. Olsen, 2006, 2007).

This article explores the contemporary use of data analytics in government and public service settings with reference to data analytics applied in smart city environments. The deployment of Big Data technologies and practices is assessed through the lens of the ‘value chain’ and the value chain conceptual model is utilized as a vehicle to aid understanding and to highlight critical issues and points of interest. It is argued that many of the Big Data challenges in modern urban environments are the result of a technocratic understanding of governance, the emergence of technologically mediated surveillance practices and conflicting practices and norms embedded in the distinction between the public and private sectors. Furthermore, the value chain approach used here highlights the different challenges that appear at different points of the chain, the differentiated role of actors in the process and how in the smart city digital sphere the public and private sectors are intimately meshed together.

Following this introduction, the next section sets out contemporary definitions of the ‘Big Data’ and ‘smart city’ concepts and seeks to align them to the research direction of the article. The subsequent section presents the value chain model and establishes its usefulness as an analytical tool for distinguishing between different stages of Big Data use. The penultimate section explores systematically each link of the value chain, as applied to Big Data in a smart city context, thereby allowing pertinent issues and challenges to be identified. In doing so, the value chain operates as a ‘roadmap’ for reviewing the extant literature. Finally, the last section offers up some concluding comments and considers the future governance of Big Data, especially in relation to smart cities where new hybrid public–private spheres of activity are emerging.

Big Data and smart cities

While analytically different, the two terms of ‘Big Data’ (or data analytics) and ‘smart cities’ are often used interchangeably for describing digital and technological investment in urban arenas (see for example, Hashem et al., 2016). As with other new technological oriented initiatives there is a tendency for discourse to focus on hype, speculation and marketing, with the aim of generating interest in commercial products and services (Nam and Pardo, 2011). This article aims to go beyond the hyperbole and to take a critical realist approach which is sensitive to governmental and public service contexts, and which highlights the ramifications of using Big Data technologies and practices in these settings. Here, the aim is to highlight the

complex intertwined relationships between Big Data technology, actors and institutions, and to highlight potential issues and consequences deriving from the use of such technologies. In this article, this is realized by using the value chain conceptual model to unpick how Big Data creates public service value and the different actors and institutions involved in this process.

Smart cities

There are notable similarities in the corporate beginnings of discourse and practices associated with Big Data and smart cities. One of the earliest users of the term ‘smart city’ is credited to IBM who have owned the trademark ‘smart cities’ since 2011 (Söderström et al., 2014). Harrison et al. (2010) note the ‘instrumented, interconnect and intelligent city’ in which various technologies, including sensors and other forms of the Internet of Things (IoT) are connected and integrated and through which data analytics are used to generate better decisions and services. Whilst this may be an example of ‘corporate story-telling’ (Söderström et al., 2014) it is how the smart city is globally portrayed and appreciated – not least among political decision-makers and service providers. The smart city promises three main areas of benefits: (1) efficient resource utilization, (2) better quality of life, and (3) higher levels of transparency and openness (Al Nuaimi et al., 2015).

Webster and Leleux (2018) argue that smart city terminology is used to capture a wide range of evolving urban practices, including new forms of service delivery, new opportunities arising from IoT and not at least new forms of governance. They posit that these technologies can enhance co-production with citizens and also open up new avenues for public participation and civic engagement. The smart city vision rests on the full utilization of information and communication technologies in general, and on data analytics more specifically. However, technology is not the only required component of a ‘true’ smart city (if such a thing exists), it also relies on social investments in urban communities, altering citizen behavior, with respect to environmental challenges and social behavior, and citizen engagement. There is a bias in the literature in which the smart city is becoming synonymous with technological development at the expense of other social and institutional challenges (Gil-Garcia et al., 2015). That is, the focus in many cities has been on technology deployment and to a lesser extent on the human and social capital (Neirotti et al., 2014).

Big Data

Although the potential benefits of data analytics have been recognized for a long time, recent technological

advancements associated with machine-learning, automated decision-making artificial intelligence and the IoT has accelerated its perceived usefulness and applicability. The term 'Big Data' was coined in the corporate environment of the 1990s and has gained currency and attention due to commercial promotions by the IT industry (Diebold, 2012). Whilst there are many definitions of Big Data, there seems to be a consensus that the term Big Data essentially captures the new possibilities of managing and analyzing large sets of both structured and unstructured data.

The academic literature on Big Data is significant and growing. However, any initial literature search reveals a bias towards technical issues, including issues relating to incompatibility and data mining, machine learning and data quality (cf. Liu et al., 2016). There is also a growing critical literature emerging from the social sciences, including from the field of public administration (cf. Dalton et al., 2016; Mergel et al., 2016; Thatcher, 2014) and urban geography (Albino et al., 2015; Kitchin 2014a, 2014b; Vanolo, 2014) which emphasizes the institutional context of technological deployment and the challenges of delivering Big Data solutions in public service and urban contexts.

Non-technical academic writing on Big Data and smart cities is still at a relatively embryonic and pre-paradigmatic stage and tends to emphasize utopian visions of what can be achieved via technological use. A more reflective critical literature is emerging on a number of issues relating to the use of data in urban environments (see for example: Kitchin, 2014a, 2014b, 2015; Mora and Deakin, 2019; Paulin, 2018; Visvizi and Lytras, 2019). Single innovative cases, mainly from the US, are often held up as indicators for wider changes that can be achieved in public policy-making and administration. Often, by the time the so-called 'corporate storytelling' of these 'innovations' has reached a broader academic audience, the actual innovations have usually already been dismantled by the organizers, or proven not to produce the anticipated desired outcomes (Söderström et al., 2014). Moreover, comparable developments in the private sector are used as both an inspiration and a benchmark for what the public sector can and should do. The rationale and functioning of the public sector, including the safeguarding of core public values, is usually ignored in exchange for the prospects of enhanced efficiency and customer-satisfaction.

Challenges to Big Data and smart cities

This article seeks to unpick what has been asserted in the literature through an analysis of Big Data based on the 'value chain' model. This perspective is inspired by

the approach taken by the International Working Group on Data Protection in Telecommunication (IWGDPT, 2014). Whilst the value chain originates in the commercial sector (Porter, 1985), there are numerous examples of public sector value chains (see for example: Davis, 2006; Heintzman and Marson, 2005) and the approach is increasingly used for the development of public service strategy (Bryson, 1988). The value chain, also referred to as value chain analysis, is a concept from business management that was popularized by Michael Porter in his 1985 best-seller, *Competitive Advantage: Creating and Sustaining Superior Performance*. The value chain categorizes the generic value-adding activities of an organization. The approach is a simple linear model that identifies a series of essential sequential activities in the creation and delivery of a good or service. In the commercial sector this approach is usually associated with satisfying consumer demand, and thereby creating value (Porter, 1985), whereas in a governmental context it can be applied to stages of policy formulation and service delivery implementation, and by doing so creating value (Davis, 2006; Heintzman and Marson, 2005). In a public service context value is realized by the service user, or consumer, and also because many public services are public goods by society more widely (Bryson, 1988).

In the commercial arena the value chain analysis enables the analyst to identify the sequence of organizational activities which contribute to the achievement of objectives. These value generating activities occur at different points in a production process. They are seen to 'add value' in that they generate financial assets, knowledge assets, information assets, expertise and skills which help the organization to achieve its ultimate goals. For Porter, this 'goal' is the endpoint of the value chain, where value is experienced, with value defined in terms of the benefits accrued from consumption of the good or service. In a public service context, the value chain analysis can also be used to identify chains of related activities that interconnect in the delivery of policy and services. This chain could involve government agencies, public service providers, private contractors and service users. Here value is multidimensional in that it is perceived to be both the value realized by the service user, the immediate consumer, and also society more broadly. In the smart city context, the beneficiary of value is further complicated by new delivery mechanisms involving public service providers and commercial partners. This raises issues about the control, ownership and access to data, and whether value is privatized by commercial interests or retained by society via public agencies. Furthermore, the complexity of the public policy and service environment also points to another way in which the value

chain analysis can be seen to be assisting with value creation. This is because many public service agencies are looking to tackle challenging ‘wicked issues’ which cut across service domains. Examples include deprivation, crime drugs and employment, all of which require multi-agency responses. Here, it is argued that there is a procedural value, which may be unquantifiable, in bringing these agencies together in a partnership scenario to collaborate in finding policy and service solutions, with the ultimate goal of realizing better services which benefit the consumer and society (Bryson, 1988). In a public service context value is therefore multidimensional, not necessarily easy to quantify, and realized by both service providers and users.

The value chain analysis can also be used to diagnose problems, blockages or failures, those organizational activities which are impeding the creation of value, in both the public and private sectors. By breaking down the production process into a series of linked activities it becomes easier to identify when one of the links in the chain is not performing adequately, creating problems or is hindering the creation of value. In essence this is how the value chain analysis is being used in this article, as a vehicle to identify the actors, organizations and activities that form the links in value creation in order to highlight specific problems that are evident in individual links, but which are less easy to identify in the broader process. The typical generic links in the value chain process are illustrated in Figure 1. Such models are intended to be simple interpretations of real life and should be seen as basic heuristic devices to help understand complex societal situations. They are particularly useful in that they can help identify different actors, actions and activities in different stages in the chain, and from there it is

possible to explore their motivations and other vested interests. In this respect, the value chain can be used as a ‘tin opener’ to unpick and understand organizational activity. Whilst the simplicity of the model can be seen as a strength it is also its undoing. Critics of the approach argue that the different stages of the model do not take place sequentially but are meshed together alongside one another, that the process is more circular than linear, that the pursuit of value underestimates other organizational forces influencing choices, and that the focus on tangible value outcomes is naïve in the modern business setting (for a critique of the value chain model see for example: Hammervoll, 2009).

Whilst not initially designed with new digital technologies in mind the value chain has over time been applied to products and services utilizing new technology. In the commercial sector variants of the value chain model with a focus on technology have emerged as the ‘virtual value chain’ (Rayport and Sviokla, 1995), the ‘innovation value chain’ (Hansen and Birkinshaw, 2007) and the ‘knowledge value chain’ (Chyi Lee and Yang, 2000). These approaches have been mirrored in the public service context with value chain approaches applied to transformations associated with eGovernment (Beynon-Davies, 2007). The approach taken in this article is to evolve the value chain model to take account of the stages of data activity associated with Big Data, especially in the smart city context. In this case the value chain extracts societal and/or individual value from processes associated with collecting, processing and using information in order to deliver better policy and services. The value chain around Big Data incorporates four sequential linked activities, from data collection/mining to storage, analysis and on to usage. Similar approaches to Big Data

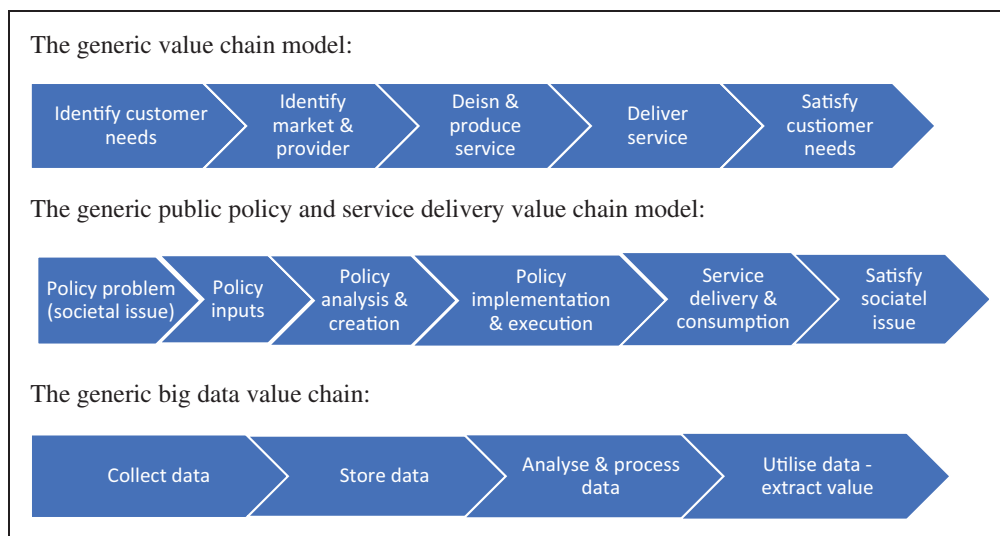


Figure 1. The value chain approach.

Table 1. The value chain of Big Data in smart cities.

Activities	1. Data collection	2. Storage	3. Analysis	4. Usage
Technologies (examples)	Social media, IoT, Sensors/cameras, 'Smart devices', Existing data sets (e.g., census)	Distributed databases 'Clouds'	Data mining Machine learning Algorithms Social network analysis Visualization	Business analytical methods Citizen profiling Predictions Open data
Variances	Voluntary, observed, inferred and coercive (legally binding)	Inhouse data storage Outsourced data storage	Data analysis Business analysis	Business analysis Direct use
Smart city examples	CCTV and other cameras, mobile device use, sensors, smart cards in mass transit, etc.	Collected and aggregated raw and unstructured data from many sources	Analytics of data from various sources	Transport management systems, Business predictions, targeting certain groups of citizens
Actors	Public sector agencies, IPs/phone companies, Social networks, Private retailers, Data vendors	Search engine providers, Social networks, Data brokers, 'clouds'	Search engine providers, Analytics companies, Government research agencies	Governments, contractors, Social Networks, Business interests, Service users, Citizens

Source: Adapted from IWGDPT (2014).

have been taken by others (see for example Davenport, 2014; Janssen et al., 2017), but not specifically in relation to the context of smart cities. The different stages of the Big Data value chain in a smart city environment, including examples of use and key actors, are presented in Table 1.

Beyond the value chain there is a broader literature on the creation of public value, including in relation to technologically mediated government services (Panagiotopoulos et al., 2019). This literature asserts that there are different types of value, or values, in different domains and that there can be a trade-off between values (Bannister and Connolly, 2014; Twizeyimana and Andersson, 2019). The public value creation literature also assesses the public value created by smart cities both in terms of improved services and cost efficiency (Meijer et al., 2016; Neumann et al., 2019). Whilst the value chain model is designed to highlight the creation of value, and in the case of Big Data in smart cities, public value, it is being used in this article as a device to identify actors and activities in the different stages of the value creation process. The model is therefore a vehicle, an analytical framework, for identifying a range of governance issues that emerge when Big Data is used for public policy and services in a smart city context.

Table 1 highlights the different technologies and actors active in each stage of the value chain. In terms of technology, the focus here are the present pre-eminent technological devices, platforms and systems, and it clear that certain technological platforms, such as the web, are prominent throughout the whole value chain. Variances in the table refer to different forms within each stage, such as the differences between volunteered, observed and coercive forms of data mining.

Examples from the smart city environment include existing projects relating to transportation (for example traffic flow/lights), public safety (for example biosecurity data), and energy and sustainability (for example supply levels of electricity and actual use). The focus on agents addresses the users and producers of Big Data and aids identification of responsibilities and vested interests. An initial interpretation of this table identifies a potential problem area in that the actors include both the public and private sectors without much consideration of the different roles they undertake, or their different motivations for being involved. A further point to note is that some of the traditional technology challenges faced by public service providers, such as paucity of technological capability, staff skills and resources, are already also prevalent in the smart city context. This includes not only access to data scientists, but also a plethora of skills and capabilities embedded in the broader 'ecosystem' encompassing new start-ups and new digital enterprises (Abellá-García et al., 2015).

The next section of the article systematically considers each step of the value chain and inherent potential challenges and problems, even if they are not overtly recognized in the dominant literature and discourse. This approach is designed to provide an analysis that goes beyond technical issues and to highlight policy, service and governance issues that arise when Big Data is used. Table 2 presents a summary of the challenges evident at each stage of the value chain.

Value chain step 1: Data collection

Quality of data. The first challenge refers to the complexity of data collection processes and how data is

Table 2. Issues associated with the value chain of Big Data in smart cities.

Activities	1. Data collection	2. Storage	3. Analysis	4. Usage
Potential issues and concerns	Quality of the data, Recycled data, Incomplete data sets, Reliability of data, Representativeness of data, Repurposed data, Data inequalities, Extent of personal data, Overt or covert collection of data, Over collection of data (mass surveillance), Legality of data	Security, Data breaches, Records management, Data destruction, Storage capacity, Data preservation, Protection of sensitive data, Data anonymization, Data protection requirements, Access and transparency, Standardization	Integrating disparate data sets, Reliability and fairness of algorithms, Acquiring data analysis skills, Interpreting data science, Satisfying commercial and public service values and logics, Integrating data analysis into public service contexts	Ownership and control of data (especially new data), IPR, Informing public policy, Data protection requirements, Informed consent, Implied or explicit anonymity, Mass surveillance, Reusing data, Fairness and transparency

constituted and used, not least who is collecting the data and for what purpose. Within a smart city context, different data sets are likely to have a number of different stakeholders involved in the collection process. First, there is data that has been voluntarily shared by individual users, consumers and citizens, as a *quid pro quo* arrangement in which access to a service is exchanged for surrendering personal data. Second, data is also collected automatically as digital transactions take place. This includes digital transactions for making purchases, the tracking of electronic building entry, vehicle and information systems, and social media activity. Third are inferred data sources where data collected for a specific purpose is recycled for an unrelated purpose. In the contemporary smart city urban environment large quantities of data are collected, including administrative, service oriented and personal data.

Some of the data collectors are likely to be agents of government, whether central or local, and subject to rigorous agreed practices for collecting data, plus implied if not explicit expectations of high methodological standards and evidence-informed procedures. However, a large group of stakeholders will be either fully commercial actors, or private contractors servicing public agencies. These actors do not need to comply with the same standards as public service providers and instead are often obliged to pursue a commercial logic. Liu et al. (2016) argue that private stakeholders prioritize profit generation and not professional standards in their data collection processes. Consequently, collection processes may not be designed for the ‘smart city public service purpose’ but for commercial purposes and may not be the result of a rigorous robust sampling process typically required in a public service context. Equally, and with special reference to social media, there are few incentives for private sector actors to value quality and fairness, and they may change both sampling and algorithmic processes if it is deemed more lucrative to do so. Indeed, social media platforms often distort data by ‘ruminating’ it as the users are making decisions based on recommendations suggested by algorithms. An example of this practice in social media is when users are suggested friends, activities and products based on automated algorithms.

Further to this, data retrieved from mobile devices in a smart city context are at best incomplete and at worst inaccurate and unreliable. Vast urban areas still suffer from poor mobile reception and the device location, particularly of mobile phones, are not directly connected to the location of the device but to the base stations of the networks. Furthermore, users often use several identities, some of which some are pseudonyms, in order to protect their privacy and to remain anonymous (Boyd, 2014; Hogan, 2013). It is

therefore not always possible to rely completely on the data retrieved from modern mobile devices even though it is increasingly used to inform public policy and service delivery in the smart city environment (Doran et al., 2016).

Finally, in relation to data collection, there is a specific challenge that derives from user, or citizen-generated, data. This relates to the vision of the smart city where the citizen is seen as an important producer of 'data', either through social media, participatory platforms, or through volunteering to generate deliberative content to online platforms (e.g., 'wikis') (Linders, 2012). A distinction can be made between social media data, where data is contributed for personal communication/networking goals and is then reused for other purposes by other actors, and 'citizen science' user-contributed data, where citizens volunteer their data for scientific and social purposes (Jennett et al., 2016). In both cases the citizen-generated data is being used to inform public policy and services (Margetts and Sutcliffe, 2013). For example, location-explicit platform applications, such as OpenStreetMaps (<https://www.openstreetmap.org>), are increasingly relevant to smart cities and are used to collect data from citizens and guide the delivery of services. However, the collection of citizen data in this way raises a number of issues, in particular with respect to the motives behind the collection where business considerations are driving the social platform affordances (Olteanu et al., 2019). Furthermore, there is a tendency for sampling bias as certain features of the data are collected more readily than others, thereby contributing to issues about the quality of the data contributed by users (Haklay, 2010). Liu et al. (2016) argue that platforms such as OpenStreetMap are biased in that they provide detailed maps for rich and wealthy regions, but only poor and incomplete information about less affluent neighborhoods.

Digital inequalities creating biases. It is generally recognized that there are inequalities in use of and access to digital technologies by citizens and service users, which in turn creates a bias in smart city data collection processes. This is evidenced in numerous studies of the users of new technology and highlights the skills required to use new digital technology, as well as access to appropriate platforms, such as access to broadband and the Internet (DiMaggio et al., 2001; DiMaggio and Harigatti, 2001). Age is often argued to be the most important demographic factor (Friemel, 2016; van Deursen and Helsper, 2015). Older age groups are often absent in attempts to retrieve representative digital data – and representation is seen as an important consideration when using such data to inform public policy. Importantly, this bias in

use does not mean the younger users' usage of technology is more valid or representative of the population. Whilst claims about 'digital natives', who are digitally literate and digitally active (Prensky, 2001) are widespread, the idea has been debunked as deeply flawed and exaggerated. Many younger users are very limited in how they use new digital technologies and lack both the skills and capacities to be considered to be fully 'native' (Bennett et al., 2008; Kirschner and De Bruyckere, 2017). Also, despite a massive proliferation of mobile devices there are still huge socio-economic inequalities in many cities, which again is reflected in certain groups' access to and use of new technology. This difference is also reflected in the use – or the 'individual production' of data – and the individual production of content, with huge biases in how users leave 'digital footprints' which are to be harvested for Big Data (Hargittai, 2007). As pointed out by Boyd and Crawford 'It is an error to assume "people" and "twitter users" are synonymous' (2012: 669). Moreover, in relation to commercial actors and their data collection methods, it is worth noting that they are usually not interested in capturing the views and behavior of less prosperous socio-economic groups as they do not represent an attractive customer group.

Privacy and consent. Although issues about privacy in a smart city Big Data context are particularly prominent at the usage stage there are also concerns relating to data collection (see for example Jain et al., 2016; Taylor et al. 2016 for a discussion of privacy issues relating to Big Data). Given the multitude of data sources, including via the IoT and social media, large quantities of data are being routinely collected without informed consent, although rules around consent change in different geographic national regions. This issue is not unique to Big Data, but it is an unavoidable element of the smart city vision. There are substantial risks associated with the 'over-collection' of data causing potential risks to individual privacy, in terms of data breaches and the (re)identification and profiling of individual citizens at a later stage in the value chain. This has been previously observed in the case of video surveillance cameras (Björklund and Svenonius, 2013; Webster, 2009). However, unlike other forms of technologically enabled practices in modern society, the contemporary privacy paradigm of 'privacy self-management' (Solove, 2013) or 'notice-and-consent' is unattainable in a smart city context given the multitude of sensory devices, multiple sources for data collection and indefinite possibilities for future data processes. So, while it is possible to notify citizens about some of the sources for data collection, for example by using 'surveillance cameras in operation' signage, it is difficult to envision a smart city urban context where people

would, or could, give informed consent to the vast array of data being collected by so many different sensory devices and platforms, not to mention the repurposing and reuse of data sets. In this respect, some of these new devices and platforms, including the IoT and the data processes they allow, appear to be at odds with basic data protection principles enshrined in European and national data protection legislation (Loideain, 2019; Maple, 2017). The risks associated with data usage are revisited in the section Value chain step 3: Data analysis.

Value chain step 2: Data storage

Security. Problems associated with data security are exaggerated by increased volumes of data and can be evidenced by the increasing frequency of large-scale data breaches (Kumar and Goyal, 2019; Singh et al., 2016). Regardless of whether data storage is centralized or decentralized the collection of large quantities of networked interlinked data makes data vulnerable and open to security challenges. Challenges relating to securing data that is resilient to tampering, malicious insiders, data loss, system failure, and data breaches is magnified in the Big Data context (Elmaghraby and Losavio, 2014). Arguably, one of the most important challenges is how to identify, isolate and protect sensitive personal information in unstructured data sets (Elmaghraby and Losavio, 2014). The widespread use of cloud computing, although not strictly a prerequisite for storing large smart city datasets, has amplified this concern. Here, an increasing amount of personal data is stored with cloud vendors in data warehouses outside the control of the organization/service which initially created the data. These clouds are often located in overseas jurisdictions and the contractors may not be subject to the same standards and professional data protection practices as the originating service provider (Lafuente, 2015). Such standards would include rules governing security protocols, data processing practices and the professionalism and integrity of the staff involved in data processing.

Access and transparency. Any discussion about access to stored data is also a discussion about the relationship between, and status of, the public and private sector organizations involved in smart city data processing (Boyd and Crawford, 2012; Johnson et al., 2017). The ultimate responsibility for collecting and storing data in smart cities is likely to be shared between public sector actors (local government, public transport agencies, utility providers, law enforcement agents, and other public service providers, etc.) and a number of private actors (social media companies, retail and hospitality industry companies, data brokers, internet providers,

cloud providers and telecoms companies, etc.). In order to realize the purported benefits of the smart city there needs to be strong relationships between these actors and efficient technical data exchange between the organizations and technological platforms and data sources involved (Johnson et al., 2017). These practices raise a number of issues – about the ability to combine the different data sets whilst ensuring data quality and integrity (Metcalf and Crawford, 2016). Furthermore, questions are being asked about whether the sharing of public service data with private companies implies that the public sector organization renounces its responsibility for future breaches and ethical standards (Yang and Maxwell, 2012) and whether the data processing norms and practices in private sector companies are as good as those in the public sector (Olteanu et al., 2019).

There are also issues about whether a private organization, whose data sets have emerged from public service sources, can deny others access to them, claim copyright and trade it to third parties. Some of these issues touch upon the global movement for ‘open government data’ in which there is an underlying assumption that increased access to government data by default is a ‘good thing’ and will provide for greater transparency, better data use and commercial opportunity. As Janssen et al. (2017) argue, there are a number of myths concerning the potential to create public value from making government data accessible. Firstly, data containing sensitive information is not suitable for general publication as it not only violates basic data protection principles, but can also be harmful for individual citizens if they are made public, for example in relation to protected identities, health, financial and lifestyle information. Furthermore, it is not clear why data collection and storage financed by the taxpayer should be given to commercial actors without charge. Many public service organizations have a legal obligation to collect data and to sell it as a revenue stream. Furthermore, if they are legally obliged to exchange that data without charge then the incentive to invest in collecting, recording and cataloguing it is compromised.

Value chain step 3: Data analysis

Incompatibility of data sets/lack of standards. At the analysis stage of the value chain there are an array of issues relating to the different technical, organizational, semantic and legal standards of the organizations and technologies involved. The core premise of Big Data is that different and often disparate data sets can be combined in new ways using algorithms and other Big Data processes to provide new insight and services. In the smart city, this can include data relating to traffic

flows, social media activity and other sensory devices, which can be used to provide a 'richer' picture of the current state of affairs. Despite this being one of the core premises of Big Data, and well-known to information system scholars (cf. Bajaj and Ram, 2007; Guijarro, 2007; Klischewski, 2004; Traumüller and Wimmer, 2004), the ability to combine diverse forms and sources of data is extremely challenging and often goes beyond the abilities of existing data integration technology (Sivarajah et al., 2017). This is not only a question of technical and semantic issues; it is also a question of organizational, political and legal differences pertaining to the use of different data sets. For example, in the smart city service context, policies (and/or inadequate or missing policies) determine the possibility of achieving joint standards in information architecture, the possibility for using algorithms and for planning for consequences and impacts (Yang and Maxwell, 2012). The 'silos' of policies and services, both vertical and horizontal, constitute the same problem for analyzing Big Data as it does for all types of digital governance (Kernaghan, 2013).

Correlation is not causation. There is a wide-spread belief that Big Data is going to revolutionize processes associated with undertaking research and predicting human behavior (González-Bailón, 2013; Janssen and Kuk, 2016). With enough data the numbers will 'speak for themselves' and if totalities can be analyzed then there is no need for theories, frameworks and models (van der Voort et al., 2019). But data does not speak for itself – 'someone' is always formulating the questions, organizing the material, conducting the analysis and interpreting the results. As expressed in the classical aphorism regarding research methods – correlation is not causation. For example, an analysis showing that there is an 80% probability that the commuters in a certain urban district will be leaving home for work between 7.30 and 8.00 am does not lead to the conclusion that all residents will be doing that. While 'traditional' research uses multiple sources to ensure research outcomes are robust and reliable there is a tendency among data analysts not to go beyond correlation (Mergel et al., 2016). The result of the Google Flu Trends analysis, as described in the seminal piece by Lazer et al. (2014), is a good example of this issue and shows how Google's attempt to detect and predict flu outbreaks (in the US) eventually failed and primarily predicted the arrival of winter. Big Data needs to be supplemented with small data (Kitchin, 2014a) and basic social science.

The power of the algorithm. Accurate prediction and the reliability of the algorithms on which they are based are

the 'holy grail' of Big Data. These algorithms have the capacity to extricate and elucidate significant patterns in future human behavior in 'real-time'. When designing an algorithm, the computer and/or data scientists and their organizations retain the privilege of formulating the purpose of the data process. For most, these algorithms are opaque and difficult to analyze or interpret (Janssen and Kuk, 2016; O'Neil, 2016). Algorithms are part of wider socio-technical assemblages (Kitchin, 2014a) and construct, reinstate and devise regimes of power and knowledge (Kushner, 2013; Steiner, 2012). In other words, the process of devising algorithms can be expected to reinforce subtle institutional bias reflecting the context in which the actors are working. This is significant for the smart city as there is an assumption that data will flow readily between public and private sector actors, but in practice the result is that the control of the design, analysis and subsequently usage, will be in the hands of commercial actors, such as social media companies, data brokers, and telecoms providers, etc., as they are main actors in data analytics. Highlighting the centrality of commercial actors in the smart city data analytics space stresses the need to respect that commercial actors are often driven by a commercial logic which is quite different to the service orientated logic of public service providers. Where a public service decides to outsource data storage and analytics it becomes possible to predict a slow drift from public service values to a more consumerist agenda reliant on and aligned with algorithms producing predictions about consumer behavior rather than citizen need (Jung, 2010). Pasquale and Bracha (2007) argue that the 'black box' of algorithms (in their case search engines), diminishes individual autonomy in the sense that the algorithms '*control informational flows in ways that shape and constrain another person's choices*' (Pasquale and Bracha, 2007: 1177). There is emerging literature and empirical research highlighting the ways in which algorithms create unanticipated and biased outcomes (Eubanks, 2018; O'Neil, 2016). This bias may not be intentional but the outcome is clearly differentiated decisions determined by computer mediated algorithms. High profile examples of algorithmic bias in the law and order arena alone include racial bias embedded in facial recognition software in policing (Janssen and Kuk, 2016), bias targeting the 'usual suspects' in predictive policing applications (Meijer and Wessels, 2019), and in algorithms assessing an individual's risk of reoffending (Andrews, 2018). The degree to which such biases exist raise governance concerns about algorithmic transparency and accountability, and fairness more generally (Binns, 2018).

Value chain step 4: Usage

Intellectual property. In the final stage of the value chain there are a number of copyright issues involved with the use of Big Data in smart city contexts (Mattioli, 2014). Many of these issues are not new, but pertain to existing legal disputes about copyrighting software, source code, algorithms and databases (Mattioli, 2014). However, there are a couple of specific issues connected to the smart city context. Firstly, the strength of data analytics in an urban context rests on the possibility to combine multiple data sets. Whilst the ownership of the raw data is normally undisputed, the question of who owns and controls the new combined and manipulated data is not always as clear. Additionally, there is an issue about whether commercial actors are able to commercialize this ‘new’ data and sell it on for a fee. This issue is aligned with the earlier discussion about open government and the prospect of the sharing of government administrative and service data with non-government actors. It is also related to the broader issue of the ownership of personal data generated by social media platforms, where the data is clearly user generated.

A reverse version of this argument is often raised in relation to the state’s surveillance capacity and whether agencies of the state have the right to trawl through personal information imbedded in commercial data sets that have been created as part of the Big Data smart city environment. Whilst most individuals appear to accept a trade-off between sharing personal data in exchange for free access to social media platforms, it remains possible to imagine some sort of redistribution mechanism for user-generated data (Smith et al., 2012). In addition to issues of ownership and intellectual property there is also the governance issue of the ‘common-pool resource’ (Prainsack, 2019) and that with ownership arises the responsibility of controlling and ensuring the accuracy of the data (Sivarajah et al., 2017).

Privacy and surveillance. Although the protection of individual privacy is pronounced throughout the value chain of Big Data, it is perhaps at the end of the value chain that the most prominent concerns arise. Proponents of Big Data often encourage the myth that because Big Data is aggregated data, and big in the sense of containing large volumes, it is therefore by default anonymized. However, re-identification is recognized as constituting a genuine threat to individual privacy (Jain et al., 2016) and not technically difficult to achieve (Ohm, 2010). This is demonstrated in the Carnegie Mellon University study, for example, where researchers managed to retrieve parts of social security number and identify individuals using

common data elements from Facebook (Stutzman et al., 2012). Skilled data analysts often only require a few data sources of personal information to be able to ‘re-identify’ individuals in Big Data sets (Jain et al., 2016; Ohm, 2010). While unique identifiers, such as full names, may be erased from the data set, there are normally other unique identifiers, such as residency, work place, age, shopping habits, etc., that are enough to identify someone.

A second set of issues concerns the recycling and reuse of data which has been collected for other purposes. A core principle in most privacy and data-protection regulation is to inform the data subject about the purpose for which the data has been collected and to secure consent when asking individuals to share information with the organization collecting data. The whole point with Big Data, and by association smart cities, is to maximize the volume of data in order to increase the value of the data set. This raises issues about how to define ‘public’ and ‘open’ data, and how consent is observed, informed and realized. Boyd and Crawford note in relation to the use of data: ‘*just because it is accessible does not make it ethical*’ (2012:671). A prime example would be the use and reuse of social media data in public service contexts. Social media data is readily available, without prior permission, because consent for its use is embedded in service user agreements. This is despite most social media data being personal data and consequently governed by data protection principles. Further to this, a smart city future with integrated urban systems combining different sources of data raises questions about who will be accountable for service failures and data breaches.

Concluding discussion

The value chain approach applied in this article illustrates that the challenges of Big Data practices in smart city contexts differ depending on the stage of the value chain. In this respect, it is a useful analytical tool which can be used to unpick data analytics in urban contexts. Importantly, issues around privacy are prominent throughout the whole process and should be central to all discussions about extending Big Data processes. Moreover, what this review also demonstrates is that the different concerns have little to do with technological maturity or technological capability. It is not just the technology that causes concern, it is the naïve belief that the technological advances are neutral and disregard political, social and economic institutions and norms.

The application of the value chain to the use of Big Data in smart city contexts highlights the intimate intertwined relationship between commercial and

public sector entities. The new digital sphere includes both hybrid data processes and the merging of values, processes and institutions. The binary public–private divide no longer exists and public service and policy outcomes are the result of the interaction of multiple organizations with competing motivations and values. The extent to which ‘value’ can be created for all parties without detriment to the other seems unlikely, given the challenges raised in this article.

Interestingly, the value chain approach highlights a range of issues and challenges at every stage or link in the process, especially where the value chain relates to a governmental or smart city context. The challenges described in this article are not meant as an argument to reject the potential of data analytics in smart cities. Rather, they point to the need for recommendations and strategies for managing the Big Data revolution in the public interest. The application of Big Data processes in smart city contexts take place in a highly fragmented space where public and private actors are assumed to collaborate and harvest value from the data, in order to deliver better public services and policy. To some extent this space is already well regulated, by data protection laws, intellectual property rights and data sharing protocols – but there are some significant gaps in relation to responsibility and accountability. Rather than advocating more formal regulation of this new sphere, where the public and private sectors are intimately meshed together, which may or may not be effective, attention could be focused on areas and practices which could be improved, which in turn could lead to more ‘value’ being extracted from Big Data whilst at the same time protecting and enhancing core public values.

First, one of the underlying themes of this review is how in a smart city context both public and private sector actors interact to constitute the foundation and infrastructure of the smart city. While many smart city applications have proven that the technical challenges can be overcome, the real issues are how to deal with organizational differences, governance and legal issues. This is not just a matter for the local actors in the smart city, but also for regional/provincial, national, and in some cases transnational, bodies of governance. Formal national and transnational regulation is likely to emerge in the longer term, but would take many years to be formed and constituted. What is needed are some basic minimum agreed self-regulatory principles concerning: (a) the quality standards of data to be used in the smart city; (b) the ethical standards regarding privacy and data protection; (c) a clearer policy regarding the ownership of unstructured and structured data; and (d) agreed standards regarding safety and protection of storing of data. These should be based on voluntary agreements between local/city

governments and main commercial actors, preferably under the auspices of some privacy ‘watchdogs’ such as information/privacy commissioners.

Second, the discourse around Big Data analytics and smart cities needs to move beyond the speculative and technological determinist space of science fiction and into a discussion about the future of urban spaces and how data analytics can be used effectively in complex urban environments. To date, most of the hyperbole around Big Data and smart cities has distorted political decisions rather than being supportive of the policy process. This is more an issue of the maturity of the discourse as opposed to the maturity of the technology itself. It took eGovernment roughly 10 years to abandon the utopian space in favor of a more ‘sober’, and perhaps more mundane, discussion about modern forms of public sector online service delivery. Here, the aspiration is that it will not take as long for discourses about Big Data and smart cities to become more critically realistic. The argument here is not that Big Data is fatally flawed, rather that it needs to be understood in its institutional environment, alongside its impacts and consequences, if individual and societal value is to be realized. Such critical realism highlights the socio-technical integration of Big Data practices and how these are operationalized in a smart city context. In other words, the true value of Big Data cannot be solely understood in technical or commercial terms and is closely aligned to other organizational and institutional practices.

Acknowledgements

The research presented in this article derives in part from the ‘*SmartGov: Smart Governance of Sustainable Cities*’ (SmartGov research project website: <http://smartgov-project.com>) research project. SmartGov is a four year (2015–2019) collaborative transnational multi-disciplinary research project on the value of ICTs for engaging citizens in the governance of sustainable cities. Funding Councils in the United Kingdom (ESRC); Netherlands (NWO), and Brazil (FAPESP) have co-funded the research. The three project partners are Utrecht University (Netherlands), University of Stirling (United Kingdom) and Fundação Getulio Vargas, Sao Paulo (Brazil).

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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The ESRC Grant reference number is ES/N011473/1.

ORCID iDs

Karl Löfgren  <https://orcid.org/0000-0002-9303-0895>

C William R Webster  <https://orcid.org/0000-0003-1211-6898>

References

- Abellá-García A, Ortiz-de-Urbina-Criado M and De-Pablos-Heredero C (2015) The ecosystem of services around smart cities: An exploratory analysis. *Procedia Computer Science* 64: 1075–1080.
- Al Nuaimi E, Al Neyadi H, Mohamed N, et al. (2015) Applications of big data to smart cities. *Journal of Internet Services and Applications* 6(25): 1–15.
- Albino V, Berardi U and Dangelico RM (2015) Smart cities: Definitions, dimensions, performance and initiatives. *Journal of Urban Technology* 22(1): 3–21.
- Andrews L (2018) Public administration, public leadership and the construction of public value in the age of the algorithm and ‘Big Data’. *Public Administration* 97: 1–15.
- Bajaj A and Ram S (2007) A comprehensive framework towards information sharing between government agencies. *International Journal of Electronic Government Research* 3(2): 29–44.
- Bennett S, Maton K and Kervin L (2008) The ‘digital natives’ debate: A critical review of the evidence. *British Journal of Educational Technology* 39(5): 775–786.
- Bannister F and Connolly R (2014) ICT, Public values and transformative government: A framework and programme for research. *Government Information Quarterly* 31(1):119–128.
- Beynon-Davies P (2007) Models for e-government. *Transforming Government: People, Process and Policy* 1(1): 7–28.
- Binns R (2018) Algorithmic accountability and public reason. *Philosophy & Technology* 31: 543–556.
- Björklund F and Svenonius O (2013) *Video Surveillance and Social Control in a Comparative Perspective*. London: Routledge.
- Boyd D (2014) *It’s Complicated: The Social Lives of Networked Teens*. New Haven: Yale University Press.
- Boyd D and Crawford K (2012) Critical questions for big data. *Information, Communication and Society* 15(5): 662–679.
- Bryson JM (1988) *Strategic Planning for Public and Nonprofit Organizations: A Guide to Strengthening and Sustaining Organizational Achievement*. San Francisco: Jossey Bass.
- Chyi Lee C and Yang J (2000) Knowledge value chain. *Journal of Management Development* 19(9): 783–794.
- Dalton C, Taylor L and Thatcher J (2016) Critical data studies: A dialog on data and space. *Big Data & Society* 3(1): 1–9.
- Davenport TH (2014) *Big Data @ Work: Dispelling the Myths, Uncovering the Opportunities*. Boston: Harvard Business Review Press.
- Davis P (2006) In search of the common wealth: A service-profit chain for the public sector. *International Journal of Productivity and Performance Management* 55(2): 163–172.
- Diebold FX (2012) A personal perspective on the origin(s) and development of “big data”: The phenomenon, the term, and the discipline (Scholarly Paper No. ID 2202843), *SSRN*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2202843 (accessed 3 March 2020).
- DiMaggio P and Hargittai E (2001) From the digital divide to digital inequality: Studying Internet use as penetration increases. In: *Center for Arts and Cultural Policy Studies Working Paper*. Princeton: Princeton University.
- DiMaggio P, Hargittai E, Neuman WR, et al. (2001) Social implications of the internet. *Annual Review of Sociology* 27: 307–336.
- Doran D, Severin K, Gokhale S, et al. (2016) Social media enabled human sensing for smart cities. *AI Communications* 29(1): 57–75.
- Ekbia H, Mattioli M, Kouper I, et al. (2015) Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology* 66(8): 1523–1545.
- Elmaghraby AS and Losavio MM (2014) Cyber security challenges in smart cities: Safety, security and privacy. *Journal of Advanced Research* 5(4): 491–497.
- Esmark A (2016) Maybe it is time to rediscover technocracy? An old framework for a new analysis of administrative reforms in the governance era. *Journal of Public Administration Research and Theory* 27(3): 501–516.
- Eubanks V (2018) Automating bias. *Scientific American* 319(5): 68–71.
- Friemel TN (2016) The digital divide has grown old: Determinants of a digital divide among seniors. *New Media & Society* 18(2): 313–331.
- Gamage P (2016) New development: Leveraging ‘big data’ analytics in the public sector. *Public Money & Management* 36(5): 385–390.
- Gil-García JR, Pardo TA and Nam T (2015) What makes a city smart? Identifying core components and proposing an integrative and comprehensive conceptualization. *Information Polity* 20(1): 61–87.
- González-Bailón S (2013) Social science in the era of big data. *Policy and Internet* 5(2): 147–160.
- Guijarro L (2007) Interoperability frameworks and enterprise architectures in e-government initiatives in Europe and the United States. *Government Information Quarterly* 24(1): 89–101.
- Haklay (2010) How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design* 37(4): 682–703.
- Hammervoll T (2009) Value creation in supply chain relationships: A critique of governance value analysis. *European Journal of Marketing* 43(5/6): 630–639.
- Hansen MT and Birkinshaw J (2007) The innovation value chain. *Harvard Business Review* 85(6): 121–130.
- Hargittai E (2007) Whose space? Differences among users and non-users of social network sites. *Journal of Computer-Mediated Communication* 13(1): 276–297.
- Harrison C, Eckman B, Hamilton R, et al. (2010) Foundations for smarter cities. *IBM Journal of Research and Development* 54(4): 1–16.

- Hashem IAT, Chang V, Anuar NB, et al. (2016) The role of big data in smart city. *International Journal of Information Management* 36(5): 748–758.
- Heintzman R and Marson B (2005) People, service and trust: Is there a public sector service value chain? *International Review of Administrative Sciences* 71(4): 549–575.
- Hogan B (2013) Pseudonyms and the rise of the real-name Web. In: Hartley J, Burgess J and Bruns A (eds) *A Companion to New Media Dynamics*. Chichester: Wiley, pp. 90–307.
- International Working Group on Data Protection in Telecommunications (IWGDPT) (2014) *Working paper on big data and privacy. Privacy principles under pressure in the age of big data analytics*, 55th meeting, 5–6 May 2014, Skopje, North Macedonia.
- Jain P, Gyanchandani M and Khare N (2016) Big data privacy: A technological perspective and review. *Journal of Big Data* 3: 25–50.
- Janssen M and Kuk G (2016) The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly* 33(3): 371–377.
- Janssen M, van der Voort H and Wahyudi A (2017) Factors influencing big data decision-making quality. *Journal of Business Research* 70: 338–345.
- Jennett C, Kloetzer L, Schneider D, et al. (2016) Motivations, learning and creativity in online citizen science. *Journal of Science Communication* 15(3): A05.
- Johnson PA, Sieber R, Scassa T, et al. (2017) The cost(s) of geospatial open data. *Transactions in GIS* 21(3): 434–445.
- Jung T (2010) Citizens, co-producers, customers, clients, captives? A critical review of consumerism and public services. *Public Management Review* 12(3): 439–446.
- Kernaghan K (2013) Changing channels: Managing channel integration and migration in public organizations. *Canadian Public Administration* 56(1): 121–141.
- Kitchin R (2014a) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.
- Kitchin R (2014b) The real-time city? Big data and smart urbanism. *GeoJournal* 79(1): 1–14.
- Kitchin R (2015) Making sense of smart cities: Addressing present shortcomings. *Cambridge Journal of Regions, Economy and Society* 8(1): 131–136.
- Kirschner AP and De Bruyckere P (2017) The myths of the digital native and the multitasker. *Teaching and Teacher Education* 67: 135–142.
- Klischewski R (2004) Information integration or process integration? How to achieve interoperability in administration? In: Traumüller R (ed) *Egov2004*. Berlin: Springer Verlag.
- Kumar R and Goyal R (2019) On cloud security requirements, threats, vulnerabilities and countermeasures: A survey. *Computer Science Review* 33: 1–48.
- Kushner S (2013) The freelance translation machine: Algorithmic culture and the invisible industry. *New Media & Society* 15(8): 1241–1258.
- Lafuente G (2015) The big data security challenge. *Network Security* 2015: 12–14.
- Lazer D, Kennedy R, King G, et al. (2014) The parable of google flu: Traps in big data analysis. *Science (New York, N.Y.)* 343: 1203–1205.
- Linders D (2012) From e-government to we-government: Defining a typology for citizen coproduction in the age of social media. *Government Information Quarterly* 29(4): 446–454.
- Liu J, Li J, Li W, et al. (2016) Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing* 115: 134–142.
- Loideain NN (2019) A port in the data-sharing storm: The GDPR and the internet of things. *Journal of Cyber Policy* 4(2): 178–196.
- Manzoor A (2015) Emerging role of big data in public sector. In Aggarwal A (ed) *Managing Big Data Integration in the Public Sector*, IGI Global. pp. 268–288.
- Maple C (2017) Security and privacy in the internet of things. *Journal of Cyber Policy* 2(2): 155–184.
- Margetts H and Sutcliffe D (2013) Addressing the policy challenges and opportunities of “big data”. *Policy & Internet* 5(2): 139–146.
- Mattioli M (2014) Disclosing big data. *Minnesota Law Review* 99(2): 535–584.
- Mayer-Schonberger V and Cukier K (2013) *Big Data: A Revolution That Will Change How we Live, Work and Think*. London: John Murray.
- Meijer A, Ramon Gil-Garcia J and Rodríguez Bolívar MP (2016) Smart city research: Contextual conditions, governance models, and public value assessment. *Social Science Computer Review* 34(6): 647–656.
- Mergel I, Rethemeyer RK and Isett K (2016) Big data in public affairs. *Public Administration Review* 76(6): 918–937.
- Metcalf J and Crawford K (2016) Where are human subjects in big data research? The emerging ethics divide. *Big Data and Society* 3(1): 1–14.
- Mora L and Deakin M (2019) *Untangling Smart Cities: From Utopian Dreams to Innovation Systems for a Technology-Enabled Urban Sustainability*. Amsterdam: Elsevier.
- Nam T and Pardo TA (2011) Conceptualizing smart city with dimensions of technology, people, and institutions. In: *Proceedings of the 12th annual international digital government research conference: Digital government innovation in challenging times*. New York: ACM, pp. 282–291.
- Neirotti P, de Marco A, Cagliano AC, et al. (2014) Current trends in smart city initiatives: Some stylised facts. *Cities* 38: 25–36.
- Neumann O, Matt C, Hitz-Gampet BS, et al. (2019) Joining forces for public value creation? Exploring collaborative innovation in smart city initiatives. *Government Information Quarterly* 46(4):101411.
- Ohm P (2010) Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review* 52: 1071–1077.
- Olsen JP (2006) Maybe it is time to rediscover bureaucracy. *Journal of Public Administration Research and Theory* 16(1): 1–24.

- Olsen JP (2007) The ups and downs of bureaucratic organization. *Annual Review of Political Science* 11: 13–37.
- Olteanu A, Castillo C, Diaz F, et al. (2019) Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data* 2(13): 1–33.
- O’Neil C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Broadway Books.
- Panagiotopoulos P, Klievink B and Cordella A (2019) Public value creation in digital government. *Government Information Quarterly* 36(4): 1–8.
- Pasquale F and Bracha O (2007) Federal Search Commission? Access, fairness and accountability in the law of search. *Cornell Law Review* 93: 1149–1191.
- Paulin A (2018) *Smart City Governance*. Amsterdam: Elsevier.
- Porter ME (1985) *Competitive Advantage: Creating and Sustaining Superior Performance*. New York: Free Press.
- Prainsack B (2019) Logged out: Ownership, exclusion and public value in the digital data and information commons. *Big Data & Society* 6(1): 1–15.
- Prensky M (2001) Digital natives, digital immigrants. *On the Horizon* 9(5): 1–6.
- Rayport JF and Sviokla JJ (1995) Exploiting the virtual value chain. *Harvard Business Review* 73: 75–85.
- Shelton T, Zook M and Wiig A (2014) The ‘actually existing smart City’. *Cambridge Journal of Regions, Economy and Society* 8(1): 13–25.
- Singh S, Jeong YS and Park JH (2016) A survey on cloud computing security: Issues, threats, and solutions. *Journal of Network and Computer Applications* 75: 200–222.
- Sivarajah U, Kamal MM, Irani Z, et al. (2017) Critical analysis of big data challenges and analytical methods. *Journal of Business Research* 70: 263–286.
- Smith M, Szongott C, Henne B, et al. (2012) Big data privacy issues in public social media. In: *Proceedings of the 6th IEEE international conference on digital ecosystems and technologies (DEST)*, Campione d’Italia, Italy.
- Söderström O, Paasche T and Klauser F (2014) Smart cities as corporate storytelling. *City* 18(3): 307–320.
- Solove J (2013) Privacy self-management and the consent paradox. *Harvard Law Review* 126(7): 1880–1903.
- Steiner C (2012) *Automate this: How algorithms took over our markets, our jobs, and the world*. New York, NY: Portfolio.
- Stutzman F, Gross R and Acquisti A (2012) Silent listeners: The evolution of privacy and disclosure on Facebook. *Journal of Privacy and Confidentiality* 4(2): 7–41.
- Taylor L, Floridi L and van der Sloot B (2016) *Group Privacy: New Challenges of Data Technologies*. Berlin: Springer.
- Thatcher J (2014) Living on fumes: Digital footprints, data fumes, and the limitations of spatial big data. *International Journal of Communication* 8: 1765–1783.
- Traumüller R and Wimmer M (2004) E-Government: The challenges ahead. In: Traumüller R (ed) *Egov2004*. Berlin: Springer Verlag.
- Twizeyimana JD and Andersson A (2019) The public value of E-government - A literature review. *Government Information Quarterly* 36(2): 167–178.
- van Deursen AJAM and Helsper EJ (2015) A nuanced understanding of internet use and non-use among the elderly. *European Journal of Communication* 30(2): 171–187.
- van der Voort H, Klievink B, Arnabold M, et al. (2019) Rationality and politics of algorithms: Will the promise of big data survive the dynamics of public decision making? *Government Information Quarterly* 36(1): 27–38.
- Visvizi A and Lytras M (2019) *Smart Cities: Issues and Challenges: Mapping Political, Social and Economic Risks and Threats*. Amsterdam: Elsevier.
- Vanolo A (2014) Smartmentality: The smart city as disciplinary strategy. *Urban Studies* 51(5): 883–898.
- Webster CWR (2009) CCTV policy in the UK: Reconsidering the evidence base. *Surveillance & Society* 6(1): 10–22.
- Webster CWR and Leleux C (2018) Smart governance: Opportunities for technologically-mediated citizen co-production. *Information Polity* 23(1): 95–110.
- Yang TM and Maxwell TA (2012) Information-sharing in public organizations: A literature review of interpersonal, intra-organizational and inter-organizational success factors. *Government Information Quarterly* 28(2): 64–175.
- Ylijoki O and Porras J (2016) Perspectives to definition of big data: A mapping study and discussion. *Journal of Innovation Management* 4(1): 69–91.