

Big data for policymaking: fad or fasttrack?

Sarah Giest¹ 

Published online: 8 August 2017

© The Author(s) 2017. This article is an open access publication

Abstract The buzz surrounding big data has taken shape in various theoretical and practical forms when it comes to policymaking. The paper combines current research streams with long-standing discussions on government and technology in public policy and public administration, such as e-government and evidence-based policymaking. The goal is to answer the question whether big data is a fleeting trend or has long-lasting effects on policymaking. Three larger themes in the literature are identified: First, the role that institutional capacity has within government to utilize big data analytics; second, government use of big data analytics in the context of digital public services; and finally, the way that big data information enters the policy cycle, focusing on substantive and procedural policy instruments. Examples from the education, crisis management, environmental and healthcare domain highlight the opportunities and challenges for each of these themes. Exploring the various aspects of big data and policymaking shows that big data is here to stay, but that its utilization by government will take time due to institutional barriers and capacity bottlenecks.

Keywords Big data · Evidence-based policymaking · Policy design · Data readiness · Digital-era governance · Policy instruments

Introduction

Big data is a broad term for the volume and complexity of data that is available. While there is no widely accepted definition of the term, the most basic description is that big data means datasets that are too large for traditional processing systems and require new technologies (Provost and Fawcett 2013). This not only refers to the size of the data, but

✉ Sarah Giest
s.n.giest@fgga.leidenuniv.nl

¹ Institute of Public Administration, Faculty of Governance and Global Affairs, Leiden University, Turfmarkt 99, 2511, DP, Den Haag, The Netherlands

also to its variety, velocity and veracity. This means that data is collected faster and that there is more variation of data that can be tapped into. Veracity refers to the uncertainty of data. This has to do both with the quality of the data, but also with the uncertainty of those dealing with the data of how accurate and complete this resource is. At the same time, the combination of digitizing administrative data, collecting data through various devices and storing more data has led to dedicated big and open data initiatives by governments. The increasingly affordable extraction of information from big data and the promise of cutting costs have also facilitated this movement. High-profile examples such as ‘data-driven campaigning’ in the 2012 and 2016 US election or the use of data for predicting where building are at risk for fires by the New York Mayor’s Office of Data Analytics (MODA) have spurred interest further. Similar developments took place in Europe, where, for example, the European Statistics Office has established a Big Data Group or the UK National Office of Statistics now has a dedicated Big Data Project.

The use of big data has been categorized as a shift at the scale of the Industrial Revolution (Richards and King 2014). Others insist that essentially nothing has changed except for datasets getting bigger. Scholars at both ends of the spectrum, however, foresee changes in the way policymaking is being done and the way it affects citizens. The former group hopes for decisions that are faster, better supported by evidence and containing less uncertainty. More critical voices revisit the obstacles outlined by the evidence-based policy discussion where different forms of information compete in the policymaking process and further require the capacity of decision-makers to comprehend it. In short, using big data for policymaking is not new, but the way the potential or actual use of big data applications changes some of the theoretical and practical discussions surrounding decision-making is.

The paper takes stock of recent theoretical developments linked to big data use in government and looks at reoccurring themes in the discussion. To illustrate the opportunities and challenges of big data, examples from various policy domains are used, such as healthcare, climate change, education and crisis management. The idea is that big data use can take shape in various forms in connection to government. There are ways to use big data for designing more effective or efficient policies, because supposedly the information decision-makers receive is more precise, vast or even predictive for the issue that they are tackling. At the same time, the regulatory framework existing in a country determines the way big data can be used by both public and private entities, because privacy laws restrict collection, sharing or utilizing personal information. Another way to think about big data and government is that of changing the manner public services are provided. This has to do with both citizens providing information to government as well as government offering personalized services based on additional data of citizens, the neighborhood or the community. Finally, the way data is dealt with within government organizations is an issue in the debate. This so-called data culture is defined as the capacity of both individual civil servants as well as the organization as a whole to collect, merge and utilize big data and the institutional structure supporting this through training civil servants or open data initiatives. Open data is the idea that data is freely available for use and reuse without ownership restrictions. These three themes, the data culture within public organizations, big and open data policy instruments and digitization of public services are looked at through the lens of current data-based theories. Those include the ‘Data Readiness Concept’ (Klievink et al. 2016) and ‘Digital-era Governance’ (DEG) (Dunleavy et al. 2005), as well as the link between big and open linked data (BOLD) as a driver of government innovation (Janssen and Kuk 2016). These theoretical concepts largely build on e-government, New Public

Management research streams and evidence-based policymaking to explain the dynamics of big data use.

The following section of the paper reviews some of the more recent data-based policymaking frameworks in order to connect them to larger research streams shaping the idea of e-government and evidence-based policymaking. The paper then addresses the three themes of data culture within public organizations (3), digitization of public services (4) and big and open data policy instruments (5). The final section (6) concludes the paper by linking the examples under each theme back to the question whether big data is a short-term trend or a long-term force changing policymaking down the line.

Data-based policymaking frameworks

Data readiness and digital-era governance

New concepts try to grasp the way the public realm is working with big data, such as the ‘Data Readiness Concept’ (Klievink et al. 2016) and ‘Digital-era Governance’ (DEG) (Dunleavy et al. 2005). Both are built on the assumption that data- and technology-driven innovations in government need an infrastructure for creating value from data and are closely linked to the e-government idea of technologies transforming government toward being more responsive and accountable (Jetzek 2016). The DEG concept is a successor of the ‘new public management’ concept and is defined as a new macro-theory for public sector development (Margetts and Dunleavy 2013). In the DEG research stream, Dunleavy et al. (2006) find that technology and digitization in public services are largely portrayed as effortless and a ‘good thing’, but that in reality, government lags behind development in the private sector, which leads to low levels of literacy connected to new technologies and at times even computers in general. This results in either government workers having to acquire a new set of skills, raising costs for personnel and training or in outsourcing of expertise. Beyond individual skills, public institutions need the capacity to ‘process information and realize desired outcomes by employing staff, creating agencies, and building up standard operating procedures’ (Dunleavy et al. 2006, 21). Generally speaking, the increased use of technology and data has had an impact on the concept and quality of information itself, where not only there is more information, but this is also tied in with a diversity of types of knowledge, as well as the capacity and skill to handle, understand and utilize it (Rose 1999; Dunleavy et al. 2006).

The data readiness concept assesses these public capacities by looking at the organizations’ data readiness and raises complementary points to DEG. The concept focuses on the organizational alignment, capabilities and maturity in connection to big data. Alignment refers to whether big data use is a good fit with the organization’s structure and main activities. Organizational maturity is the maturity of e-government initiatives within this organization, and finally, the capabilities describe the organization’s use of big data linked to IT and data governance, data science expertise or legal compliance. The concept connects these characteristics to a value chain in the big data process that includes collection, combination, analysis and use of data (Klievink et al. 2016). Especially the organizational maturity criterion is rooted in the e-government tradition of looking at the e-government growth stages: stovepipe organizations, integrated organizations, nationwide portal, inter-organizational integration and demand-driven, joined-up government (Klievink and Jansen 2009). In short, the more advanced the public organization is in terms of adapting to

environmental changes, the better its performance when it comes to digital government infrastructures (Klievink and Janssen 2009).

Janssen and Kuk (2016) go one step further and identify big and open linked data (BOLD) as a driver of government innovation. Janssen et al. (2017) hypothesize that in the ecosystem of private actors and citizens, government is under pressure to adapt its institutional structures to new forms of data that then affect the delivery of policies. This data-driven innovation is facilitated by factors that can be strategic and political, organizational, linked to data governance or purely technical. A distinction among more general public innovation and data-driven innovation in the public sector is that it is not necessarily driven by public organizations, but might be facilitated by private organizations as well as citizens, which can result in new organizational forms (Janssen et al. 2017):

Old government structures need to be changed and a shift from inward-looking towards outwards-looking is necessary. Trust among parties is a prerequisite to make this work. Policies providing incentives for collaboration and to organize collaboration between public and private actors can drive this kind of innovation. (Ibid, 191)

This can lead to four different data-driven innovation types:

1. Co-creation-based innovation
2. Crowdsourcing-based innovation
3. Service innovation
4. Policymaking innovation.

These categories contain different levels of public and private involvement as well as the usage of the information produced (Janssen et al. 2017). Co-creation- and crowdsourcing-based innovation both have high levels of participation and have an external component, as information is gathered outside of government with the goal of giving input on public developments. In contrast, the innovation in service is more indirectly affected, as private companies might develop services for citizens based on open data that government has to compete with or incorporate, which can result in public service innovation. Policymaking innovation is the idea that government can use data to model future policy implications and support potential policy decisions (Janssen et al. 2017). These theoretical developments closely mimic the discussion in the e-government field, where Bannister (2001) distinguishes among three themes, (1) the improvement and execution of public services linked to new technologies, (2) technologies transforming the way government is organized and (3) technologies boosting values such as transparency and accountability.

Evidence-based policymaking

These rather recent trends in the literature tie in with the broader concept of evidence-based policymaking. The evidence-based policymaking research largely evolves around the receptiveness of the policy development ‘cycle’ toward such input, and there is a debate on where and how evidence-based contributions can add value to the process (Head 2008). Two aspects that have been raised in this context are the integration of big data in an existing institutional context and the capacity of individuals or government entities to be able to find and utilize data-based information. Both are connected in that limited capacity available within government can lead to the involvement of additional actors, which ultimately increases the level of institutional complexity. The capacity is defined as ‘political analytical capacity,’ which describes that when governments experience low levels of analytical capacity, they risk incorporating scientific knowledge ineffectively into the

decision-making process (Sanderson 2006; Pawson 2006; Nutley et al. 2007; Howlett 2015). This often results in adding additional stakeholders that possess the skills needed to extract relevant information from the given source. Additional stakeholders add to the institutional complexity and can shape the way policy and evidence interact. Best and Holmes (2010) point out that coordination across several departments becomes increasingly difficult as complexity grows, which can lead to ineffective inclusion of evidence. It further slows down policy processes, because beyond purely technical input, other information sources are consulted (Sanderson 2006). Several scholars suggest that compromises among political and technocratic elements are made in this process, where the rational idea of using robust evidence is mixed with political ideology and other ‘non-evidence-based’ ideas (Best and Holmes 2010; Howlett 2009).

Lavertu (2014) further links big data to evaluating policy processes. He warns of the ‘imbalances in the precision of performance metrics’ that might, in a second step, lead to goal displacement in organizations that deliver public programs (Holmstrom and Milgrom 1991; Lavertu 2014, 866). This has to do with two issues, first, the failure to connect performance measure data with outcome dimensions, leading to inaccurate findings regarding the contribution of public organizations toward certain societal outcomes. Second, aggregated performance measures, such as school or teacher performances, are often publicly available, which poses an opportunity for external stakeholders to affect policy-making down the line based on voting behavior or lobbying (Lavertu 2014).

Policy design lens

At a more abstract level, these processes can be described through the lens of policy design. This concept is linked to the idea that governments aim to implement goals effectively and efficiently, and connected to that, are interested in utilizing knowledge and experience about policy issues (deLeon 1999; Howlett 2011). In the formulation stage, policymakers define policy options. This is where much of the design activities come into play, but can also reach beyond formulation by representing ideas that might re-occur in practice (Goggin 1987; Howlett 2011). The policy design concept looks at these considerations in policy formulation and the outcomes in implementation. This perspective pays special attention to policy instruments, which are defined as ‘the toolbox from which governments must choose in building or creating public policies’ (Howlett 2011, 22). Thereby, the selection of policy instruments takes place within a larger context that contains institutions, actors and practices and that affect the policymaking process.

Linking these ideas to big data, information-based implementation tools highlight some of the variations in using data for pursuing certain policy outcomes. Howlett (2011) distinguishes between substantive and procedural informational instruments, which are connected to different aspects of policymaking. Substantive information collection and dissemination tools describe government collecting information to enhance evidence-based policymaking, and public institutions communicating information to citizens through, for example, information campaigns. Procedural information tools describe the activities by government to regulate information based on information legislation for the release of, for example, government data.

Taken together, these approaches to data-based policymaking carry different labels, but they converge on several themes. First, the idea that government entities require the capacity, skills and data culture to deal with this type of evidence. Second, the notion that this data is used by government to engage citizens and digitize public services. Finally, the role of big data in policymaking, where government uses various information policy

instruments for reaching policy goals. These themes will be addressed in more detail below.

Data culture within public organizations

Data culture within public organizations refers to understanding big data not only as an IT issue, but as something that requires support from organization-wide structures and capabilities (Helfat et al. 2007; Comuzzi and Patel 2016). Specifically, it emphasizes the importance of civil servants and policymakers understanding how to find, analyze and utilize big data and the institutional structure to support this through, for example, training or sharing of data among government departments. This is something that is mentioned in passing in much of the recent literature on big data and public policy. For example, the Data Readiness Concept references this in the organizational capabilities category, where Klievink et al. (2016) identify IT governance, IT resources, internal attitude, external attitude, legal compliance, data governance and data science expertise as relevant factors for dealing with big data. Similarly, the DEG framework incorporates the idea that the levels of literacy connected to new technologies within government are often low and government workers have to be trained in new skills (Margetts and Dunleavy 2013). The evidence-based policymaking framework also addresses this in form of ‘political analytical capacity’ of governments (Sanderson 2006; Pawson 2006; Nutley et al. 2007; Howlett 2009).

It follows that there are two more general aspects that play into the use of big data in government. First, IT systems within government are dependent on institutional mechanisms enabling their development (Dunleavy et al. 2006). Subsequently, the weaker this mechanism is, the less successful the IT performance. Second, limited capacity within government to utilize big data can have effects on who handles the data. Much of the IT used within government is outsourced and being delivered by private stakeholders, which increases the impact of industry on changes and performance of IT within government (Dunleavy et al. 2006).

Institutional context of data use

Potentially weak institutional mechanisms can be traced back to several factors, some of which will be outlined below with examples. The first one is the siloed (data) structure that many government departments encounter. This includes a legal component where data cannot be shared due to privacy laws, but more often than not it is the institutional setup and routine of sharing and collecting data that poses a major obstacle. Whereas public silo systems were associated with the institutional structure, they now also include an IT and data element (Bannister 2010). The IT silo systems (also called stovepipe systems) describe a system which was developed to reduce complexity and create clear rules of reporting and decision-making; however, due to increased collaboration and interdepartmental topics, these have become obstacles in the policymaking process.

Electronic healthcare (ehealth) is a widely used example in this scenario (Nedlund and Garpenby 2014). This has to do with the fact that each new medical device has its own database, which can create a new silo system (Bygstad et al. 2015). In the case of the UK NHS electronic patient records for example, the digital transition was a way to tackle the departmental and organizational silos of patient data, as most records were paper-based,

which made information sharing time-consuming and inefficient. At the same time, it was seen as an opportunity to run big data analytics for resource allocation and healthcare initiatives. The digitization has, however, led to data format inconsistencies where data is stored in diverse ways and formats and data on drugs, staff or locations are recorded differently. This makes it difficult to share the relevant data or reach meaningful conclusions (Ford 2016). In a recent report, the sharing and integration of data remains an issue, as it is highlighted that ‘improvements must be made to the ease and safety of sharing data between services’ (Care Quality Commission 2016). Along the same lines, the report also points toward a gap in training staff to handle data safely and share it across departments (Care Quality Commission 2016).

Another case where the institutional structures limit the use of data is carbon emission reductions at local level. Several UK municipalities participated in the Department of Energy and Climate Change’s (DECC) Local Carbon Framework (LCF) program. This framework will serve as a local action plan on delivering carbon emissions, encapsulating the varying portfolios of carbon reduction measures relevant to individual or grouped councils (Gray et al. 2011). In this context, the municipalities heavily rely on national and interdepartmental data to assess carbon emissions and potential policy initiatives. The evaluation report for the implementation of data-driven initiatives highlights several issues linked to data use (Giest 2017). The Dorset Energy Group points toward ‘limitations of national data’ while the Manchester group highlights the ‘lack of national consistency and standardization’ and Bristol talks about data that was ‘out of date’ (Gray et al. 2011, 203–204).

Outsourcing of data services

Another factor that contributes to weak institutional mechanisms for technological and data development in government is institutional complexity. This is partially addressed by the stove-piped structure that prevents civil servants from sharing information, but it also incorporates the idea that stakeholders are added to the policymaking process, which can have an effect on decision-making procedures. The use of big data analytics requires more privatization and contracting out of government activities linked to accessing, combining and making sense of data as well as collaboration across departments and within communities (Bătăgan 2011; Meijer and Bolivar 2015). This is driven by limited expertise within government to deal with the data and often leads to public officials working with stakeholders that they have no experience with (Radin 2003). A study by Ernst and Young (EY) (2013) finds that in Northern Europe (Sweden, Norway and the UK) access to specific knowledge, expertise and tools are key drivers for outsourcing rather than cost-efficiency. A similar trend is emerging in the USA where a 2014 survey indicates that:

While the federal workforce is increasingly basing decisions on data, it lacks the data and analytics skills to translate complex datasets into useful knowledge for decision-makers. Most respondents (78 percent) called data a significant component of their jobs, and 60 percent scored their use of data to make decisions above average. But a stunning 96 percent identified a data skills gap at their agency (SAS 2014, 1).

There is further evidence that such differentiation and specialization aggravate coordination issues in government. Government officials in the UK raise concerns that private stakeholders neglect how a privately developed technology can be integrated into the municipal environment and might lead to limited information flow among partners and

government departments down the line (McKinsey 2014). Another issue raised in this context is that using different IT systems within the same government might reinforce existing institutional silos (Copeland 2014).

This more inward-looking perspective on government and the use of big data shows that incorporating big data information into administrative and policymaking routines challenges existing structures and applications. Updating these takes time or the processes are outsourced, which comes with a new set of challenges.

Digitization of public services

The digitization of public services is closely linked to the use of data: Digital applications provide ample opportunity to aggregate and analyze data, and, in turn, data analysis can support the implementation of digital services (Demirkan and Delen 2013). This translates, for example, into open or personalized data where government collects citizen data and then provides a service based on this data that makes accessing and using public services easier. Another application is that of citizen giving additional data or even collecting data for government, supporting public service delivery down the line.

In theoretical terms, digital public services (e-services) are largely looked at from an e-government perspective. The DEG framework addresses this aspect by looking at how the digitization of administrative processes progresses (Dunleavy et al. 2005). Digitization is further coupled with the idea of defragmentation or centralization of structures to standardize the technology, methods and data being used (Rose and Grant 2010) as well as opportunities for citizen input and collaboration. This includes creating services that are more responsive to the needs of citizens as well as government itself being more efficient in its response (Bekkers and Homburg 2007). Citizen participation is also linked to a push for open data since the assumption here is that open data can lead to increased transparency and accountability regarding public entities and services and could potentially promote public participation in decision-making (Yiu 2012). The design of digitization includes both decisions on the format and type of technology as well as the organizational structure connected to it. For digital service implementation, more detailed aspects, such as training users, data conversion or systems maintenance activities are involved (Melin et al. 2016). Therefore, data and data information form a subset of factors playing into the development and execution of digital services. This also covers, for example, ‘the capture, management, use, dissemination, and sharing of information’ and data quality and accuracy aspects (Melin et al. 2016, 13; Gil-García and Pardo 2005). As the examples will show, these data-related aspects are further complicated within a multi-actor arrangement inside government that is restrained by the national and local institutional context (Wesselink et al. 2014).

E-services

An example for government collecting citizen data and providing a public service in return is an online citizen portal. Borger.dk is the Danish Citizen Portal for accessing personalized data and services through a digital signature.

Based on personalized data maintained by public authorities, citizens can access personalized services such as data about economy, e.g. salary received for the last three months; taxes paid; housing, e.g. property value or location; and civil registry data, e.g. social security number or children’s and spouse’s social security number. In

addition, a link is provided to update personal data and print relevant documents. (Bertot et al. 2016, 218)

This service was launched in 2010 and, according to the OECD (2014) developed by all government levels (central, regional, municipal). The portal is operated by the Danish Agency for Digitization, within the Ministry of Finance (OECD 2014). It is the central service point for citizens to reach the Danish government online and is part of a larger European trend to integrate various digital services in one platform, such as the UK's gov.uk or the Dutch mijnoverheid.nl. These changes require government departments and back offices to exchange information and make data accessible. This goes hand-in-hand with continued political support and citizen engagement strategies for the proper uptake of the technology among government officials and citizens. Research in the Australian context shows that a dedicated policy linked to citizen portals, such as open data regulations and resources, enhances its quality and, in a second step, citizen uptake. It further appears that outsourcing such applications either to the state or the federal government level is not positively associated with service capabilities over time (Chatfield and Reddick 2017).

Crowdsourcing

Another example, which falls into the second category of citizens providing additional data, is the Disaster Reporter App by the US Federal Emergency Management Agency (FEMA). The app allows users to upload photographs and send short texts about a disaster region. At the same time, survivors can access information and maps through the app during and after a crisis. FEMA officials further gain insight into the affected region and, based on this information, make decisions on resources needed and emergency routes to take. 'Digital tools for situational awareness include social media, GIS, sensors, big data, bio-data, and environmental data, as well as analytical algorithms, prediction and outcome modeling, and tools to assist in decision-making, resource allocation, and response strategies' (Bertot et al. 2016, 218). This crowdsourcing approach to a crisis allows decision-makers to use real-time information coupled with existing emergency plans. Critics, however, point toward two issues that arise in this scenario: First, government agencies could easily get overwhelmed by the information influx during a crisis and end up being unable to sort and utilize all the relevant information. This is especially the case when there is no automated support, e.g., an algorithm that can sort pictures and information based on their geo-tagging or could source additional information from social media services, such as Twitter. Civil servants also require proper training to deal with both the new technologies and the influx of data. Second, in contrast to the first point, crowdsourcing information only works when enough people use the app to submit pictures and provide real-time data. In other words, the app needs to be demand-driven for enough citizens to download it and actively use it during a crisis situation (Meier 2013).

More generally, the use of crowdsourcing for decision-making is contested. The basic idea is that decision-making would be more decentralized if the 'wisdom of the crowds' is used to gather diverse information paired with a push for change in contrast to established stakeholders who are prone to argue for the status quo (Surowiecki 2004; Lodge and Wegrich 2014). The hope is that crowdsourced information can increase regulatory quality by giving governmental stakeholders the opportunity to use public opinion as a way of changing current legislation in an environment that is inexpensive due to utilizing technology. It further poses a direct link between the concerns citizens, companies or other

stakeholders raise and policymakers (Lodge and Wegrich 2014). Digitizing this process is further linked to the expectation that decision-making is be more transparent and open to suggestions and potential consequences would be revealed earlier in the process. This could ultimately lead to more informed decision-making. These rather idealistic notions have been challenged in the past, and the criticism largely focuses on the limited impact digital crowdsourcing has had on, for example, regulatory changes, due to pre-defined administrative procedures and the limited flexibility on government-side to incorporate additional or new information (Beierle 2003; Lodge and Wegrich 2014). This brings the argument back to the idea that the institutional setting and the processes within government limit the impact that digital technologies and the data collected with it can have on decision-making.

Big and open data policy instruments

Connecting data culture within government and digital services to policymaking shows that increased data use has swept through many policy areas and shaped procedural and substantive policy instruments. Thereby, government is both data producer and consumer. The vast amount of administrative data collected at various governmental levels and in different domains, such as tax systems, social programs, health records and the like, can—with their digitization—be used for decision-making in areas of education, economics, health and social policy. In addition to these more traditional data, governments and companies increasingly add more (real-time) data based on social media input, cameras and sensors. Recent work highlights this by using administrative data for applying novel research designs and linking records to track outcomes of experiments and quasi-experiments (Einav and Levin 2014). Rather than evaluating average policy treatment effects, studies ‘build models that map individual characteristics into individual treatment effects and allow for an analysis of more tailored and customized policies’ (Einav and Levin 2014, 719).

Well-known examples come from the healthcare sector. Personalized medicine, such as individualized diagnosis and treatments, are delivered based on data. Clinical decision support systems have been facilitated by the automated analysis of X-rays or computed tomography (CT) scan images. Finally, the reliance on patient-generated data is increasing based on the use of mobile devices of citizens and sending educational messages for behavioral changes (Roski et al. 2014). These developments have the goal of saving costs and supporting the standardization of care (Murdoch and Detsky 2013). Administrative data have been similarly useful in documenting regional disparities in economic mobility.

Researchers have used large-scale administrative data to measure and compare relevant variables (e.g., income, spending, productivity, or wages) across small sub-populations... These results have helped guide policy discussions and define research agendas in multiple subfields of economics. (Einav and Levin 2014, 717)

These data analyses have not only been used to track policy implementation, but also to offer insight into agenda setting processes through adding social media platform data (Neuman et al. 2014). Using big data for information-based policies can further be linked to substantive and procedural policy instruments.

Procedural policy instruments

Procedural information tools describe the regulatory activities by government, which include:

Creating (and destroying) demand for various technologies through regulation; conducting and supporting R&D activities in support of environmental goals; promoting technologies through subsidy; and facilitating knowledge transfer between government, regulated firms, and outside environmental equipment suppliers through everything from the patent system to industry-specific conferences, publications, and collaborations. (Taylor et al. 2005, 348–9)

An example for these dynamics is the policy framework for big and open data. Open data is data that is detached from copyright, patents, censorship and similar restrictions linked to data dissemination, which has increasingly incorporated big data. The idea behind sharing government data is that citizens, companies and nonprofit organizations have the opportunity to use it for new products or services (Bertot et al. 2014). ‘Big and Open Data initiatives have the potential to lead to new scientific and research insights, create economic development, inform decision and policymaking and generate new policies that benefit the publics served by government’ (Bertot et al. 2014, 6). These goals, however, also pose a major challenge to government. Based on the example of the Open Government Directive (OGD) in the USA, several authors highlight that the lack of specific guidelines hindered federal agencies’ implementation efforts—specifically the introduction of data.gov (McDermott 2010; Evans and Campos 2013; Bertot et al. 2014). Looking at the information management documents by the US Office of Management and Budget (OMB), Bertot et al. (2014) emphasize that the policies ‘provide broad principles and guidance for agencies, but fail to address the use of Big and Open Data, as nearly all pre-date the development and use of Big Data Technologies’ (Ibid, 10). This results in barriers for agencies using and distributing data concerning its quality, labeling of datasets, confidentiality and privacy. A more concrete example is that of the US education agency in connection to the Open Government Directive (OGD), which had to set up a Department-wide Disclosure Review Board responsible for coordinating, reviewing, and approving the privacy of public data releases, develop a plan for technical assistance to state and districts on the subject of privacy protections as well as offer targeted assistance for education stakeholders upon request (Department of Education 2016).

Substantive policy instruments

Substantive information tools describe government collecting data to enhance evidence-based policymaking. The education sector is one example for this, as public educational institutions increasingly utilize big data for creating digital and interactive data visualizations that can then give up-to-date information on the education system for policy-makers. This further involves a predictive element, where trajectories of individual students can be created and calculated for future performance of both the system and the learner. Educational policy has increasingly focused on these types of knowledge sources (Edwards 2014; Williamson 2016):

Learning analytics constitutes an emerging form of policy instrumentation in educational governance privileging techniques of prediction and preemption. Such ‘big data’ practices are distinct from the large-scale datasets used in contemporary

techniques of government (such as international assessments). (Williamson 2016, 125)

The idea is that those feedback moments for government are available synchronously and automatically to allow information for both short-term and long-term decision-making processes in the education sector, such as adjusting policy instruments or publishing new guidelines for public schools. But similar to the healthcare examples, schools struggle to store, process and provide access to the data. The data is stored in various databases and often has incompatible formats or requires different passwords. This makes gathering the data time-consuming and less attractive for schools to provide comprehensive data-based information. It has also led to schools involving companies from the ‘EdTech’ industry to provide big data techniques (Charlton et al. 2013; Carmel 2016). Such techniques include learning analytics (LA) and educational data mining (EDM). Both have the ability to identify patterns in the data, conduct fine-grained analysis over long periods of time and analyze the effects of learning environments on students (Baker 2013). In addition, some have raised concern that policy decision-making based on the data mining models will exacerbate bias and create new forms of discrimination, due to algorithms that reflect norms and values and could reinforce structural inequalities and cumulative disadvantages (Alarcon et al. 2014; Carmel 2016).

This was an issue in the so-called IMPACT program in Washington DC. The data-based evaluation tool for teachers spans nine performance criteria covering clear presentation, behavior management and skills. Teachers are graded on a one to four scale (ineffective, minimally effective, effective and highly effective). This is paired with human observation in the classroom. Being rated ‘ineffective’ twice results in a termination of the contract. After pressure from education reformers for more quantifiable and rigorous ways to evaluate teachers, this system was installed in the USA (McCrummen 2011). Critics, such as unions, researchers and educators, however, point out that test results are too vulnerable to conditions outside a teacher’s control, such as poverty, learning disabilities and random testing day incidents such as illness, crime or a family emergency that can skew scores (Turque 2012). Since then, the algorithm has been updated and training programs have been set up to increase evaluator reliability (Gitomer et al. 2013).

Concluding remarks

The foregoing overview gives an idea of how big data is situated in the policy field—both theoretically and practically. The article takes a broad perspective on big data trends rooted in the Public Administration and Public Policy literature to point to future directions for research and give examples from different policy domains, such as health, education, climate change and crisis management. The concepts of Digital-era Governance (DEG), Data Readiness, Evidence-based Policymaking and Policy Design all link directly to public use of new technologies and big data streams. More importantly, while each theoretical perspective emphasizes different opportunities and challenges in junction with new data developments, they converge on two main aspects: First, the fact that existing administrative and institutional structures define the way data is collected, analyzed and used due to limited institutional support and data silos. And second, the capacity within government plays a role in how data is dealt with or used at all when looking at specific policy domains. The diverse examples given under the themes of public data culture, digitization of public services, and big and open data policy instruments all come back to these two challenges in

one form or another. The examples further highlight that problems occur in different dimensions of big data—not only in the use of data to tackle issues, but also in how the information that this data contains enters the policy process and ultimately affects policy decisions as well as regulatory frameworks directing data collection efforts and sharing of information.

To conclude, many of the issues raised in the context of public big data use are not new, as they have been addressed in waves throughout the history of government incorporating technology and digital services into administrative processes and various policy domains. The big data movement, however, has moved past the question of ‘if’ and is much more about the ‘how’: How can big data be incorporated into policymaking at different governmental levels, how can big data be regulated and how can it be utilized? Drilling down on each of these would reveal a complex picture that offers a variety of research streams still to be explored, but the current assessment shows that there is no turning back from the expectation that more data can lead to more information and eventually a more efficient and effective government. On the surface, the challenges arising are old, such as institutional context and policy capacity limiting big data use; however, upon closer inspection, the issues are more nuanced in that there is a digital component to the physical structures and processes that requires attention. These digital elements add specific expertise, personnel and technology to the mix and, depending on the issue at hand, might first complicate public policymaking before offering potential efficiency and effectiveness down the line. In short, while big data is not a fad, it also is not a fast track in the early phases of its application.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Alarcon, A., Zeide, E., Rosenblat, A., Wikelius, K. Boyd, D., Gangadharan, S. P., and C. Yu (2014). Data and civil rights: Education primer. Data and Civil Rights Conference, October 30, 2014 . http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2542268. Accessed April 30, 2017.
- Baker, R. (2013). Educational Data Mining and Learning Analytics. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (2nd ed.). Cambridge: Cambridge University Press.
- Bannister, F. (2001). Dismantling the silos: Extracting new value from IT investments in public administration. *Information Systems Journal*, 11, 65–84.
- Bannister, F. (2010). Deep E-Government, Beneath the Carapace. In H. J. Scholl (Ed.), *E-Government, information, technology and transformation* (pp. 33–52). London and New York: Routledge.
- Bătăgan, L. (2011). Smart cities and sustainability models. *Informatica Economică*, 15(3), 80–87.
- Beierle, T.C. (2003). *Discussing the Rules: Electronic Rulemaking and Democratic Deliberation*. Discussion Paper No. 03–22. Resources for the Future, Washington, DC.
- Bekkers, V., & Homburg, V. (2007). The Myths of E-Government. *The Information Society*, 23(5), 373–382.
- Bertot, J., Estevez, E., & Janowski, T. (2016). Editorial, Universal and contextualized public services: Digital public service innovation framework. *Government Information Quarterly*, 33, 211–222.
- Bertot, J., Gorham, U., Jaeger, P., Sarin, L., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information Polity*, 19, 5–16.
- Best, A., & Holmes, B. (2010). Systems thinking, knowledge and action: Towards better models and methods. *Evidence and Policy*, 6(2), 145–159.
- Bygstad, B., Hanseth, O., and S. Truong Le. (2015). *From IT silos to integrated solutions*. A study in ehealth complexity. ECIS 2015 completed Research papers. Paper 23. ISBN 978-3-00-050284-2.

- Care Quality Commission (2016). Safe data, safe care. Newcastle UK, July 2016, CQC–304–072016.
- Carmel, Y. H. (2016). Regulating “big data education” in Europe: Lessons learned from the US. *Internet Policy Review*. doi:10.14763/2016.1.402.
- Charlton, P., Mavrikis, M., and D. Katsififi (2013). The potential of learning analytics and big data. Ariadne, 71. <http://www.ariadne.ac.uk/issue71/charlton-etalthash.wainfh00.dpuf>. Accessed April 25, 2017.
- Chatfield, A. K., & Reddick, C. G. (2017). A longitudinal cross-sector analysis of open data portal service capability: The case of Australian local governments. *Government Information Quarterly*. doi:10.1016/j.giq.2017.02.004.
- Comuzzi, M., & Patel, A. (2016). How organisations leverage big data: A maturity model. *Industrial Management and Data Systems*, 116(8), 1468–1492.
- Copeland, E. (2014). Small pieces loosely joined, how smarter use of technology and data can deliver real reform of Local Government. *Policy Exchange*. <http://www.policyexchange.org.uk/webreports/small-pieces-loosely-joined/>. Accessed April 25, 2017.
- deLeon, P. (1999). The stages approach to the policy process: What has it done? Where is it going? In P. A. Sabatier (Ed.), *Theories of the Policy Process* (pp. 19–34). Westview: Boulder, CO.
- Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decisions Support Systems*, 55, 412–421.
- Department of Education (USA) (2016). Open Government Plan, September 2016 Update. <https://www.state.gov/documents/organization/262834.pdf>. Accessed April 30, 2017.
- Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2005). New public management is dead. Long live digital-era governance. *Journal of Public Administration Research and Theory*, 16, 467–494.
- Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2006). *Digital Era Governance: IT Corporations, the State, and E-Government*. New York: Oxford University Press.
- Edwards, R. (2014). Software and the hidden curriculum of digital education. *Pedagogy, Culture and Society*. doi:10.1080/14681366.2014.977809.
- Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science*, 346(6210), 715–721.
- Ernst and Young (EY) (2013). Outsourcing in Europe: An in-depth review of drivers, risks and trends in the European outsourcing market. http://www.ey.com/Publication/vwLUAssets/Outsourcing_in_Europe_2013/%24FILE/EY-outsourcing-survey.pdf. Accessed April 30, 2017.
- Evans, A. M., & Campos, A. (2013). Open government initiatives: Challenges of citizen participation. *Journal of Policy Analysis and Management*, 32(1), 172–185.
- Ford, S. (2016). Breaking down EMR Silos for Paperless NHS. Accenture Post, December 12th, 2016. <https://www.accenture.com/gb-en/blogs/blogs-breaking-down-emr-silos-paperless-nhs>. Accessed April 25, 2017.
- Giest, S. (2017). Big data analytics for reducing carbon emission in smart cities: opportunities and challenges. *European Planning Studies*, 25(6), 941–957.
- Gil-García, J. R., & Pardo, T. A. (2005). E-government success factors: Mapping practical tools to theoretical foundations. *Government Information Quarterly*, 22, 187–216.
- Gitomer, D., Crouse, K., and J. Joyce. (2013). A Review of the DC IMPACT Teacher Evaluation System. Center for Education Policy Analysis, October 21, 2013.
- Goggin, M. (1987). *Policy design and the politics of implementation: The case of Child Health Care in the American States*. Knoxville: University of Tennessee Press.
- Gray, D., Couper, G., Jones, E., Wade, J., Pilling, A., Machin, N., & Matthews, P. (2011). Evaluation of the local carbon framework pilots. A report by CAG Consultants in association with Impetus Consulting and Dr Joanne Wade, Commissioned by the Local Government Association and the Department for Energy and Climate Change. Retrieved May 8, 2017, from http://www.local.gov.uk/c/document_library/get_file?uuid=5c86ec51-3e42-4efc-bce4-35cfa7f0b9ac&groupId=10180.
- Head, B. (2008). Three lenses of evidence-based policy. *Australian Journal of Public Administration*, 67(1), 1–11.
- Helfat, C. E., Finkelstein, S., Mitchell, W., Peteraf, M., Singh, H., Teece, D., et al. (2007). *Dynamic capabilities: Understanding strategic change in organizations*. Carlton: Blackwell.
- Holmstrom, B., & Milgrom, P. (1991). multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law Economics and Organization*, 7, 24–52.
- Howlett, M. (2009). Policy analytical capacity and evidence-based policy-making: Lessons from Canada. *Canadian Public Administration*, 52(2), 153–175.
- Howlett, M. (2011). *Designing public policies, principles and instruments*. New York: Routledge.
- Howlett, M. (2015). Policy analytical capacity: The supply and demand for policy analysis in government. *Policy and Society*, 34, 173–182.
- Howlett, M., Ramesh, M., & Perl, A. (2009). *Studying Public Policy, Policy Cycles and Policy Subsystems*. Oxford: Oxford University Press.

- Janssen, M., Konopnicki, D., Snowdon, J., & Ojo, A. (2017). Driving public sector innovation using big and open linked data (BOLD). *Information Systems Frontiers*, 19(2), 189–195.
- Janssen, M., & Kuk, G. (2016). Big and open linked data (BOLD) in government: A challenge to transparency and privacy? *Government Information Quarterly*, 32, 363–368.
- Jetzek, T. (2016). Managing complexity across multiple dimensions of liquid open data: The case of the Danish basic data program. *Government Information Quarterly*, 33(1), 89–104.
- Klievink, B., & Janssen, M. (2009). Realizing joined-up government: Dynamic capabilities and stage models for transformation. *Government Information Quarterly*, 26(2), 275–284.
- Klievink, B., Romijn, B., Cunningham, S., & de Bruijn, H. (2016). Big data in the public sector: Uncertainties and readiness. *Information Systems Frontiers*, 19(2), 267–283.
- Lavertu, S. (2014). We all need help: ‘Big Data’ and the mismeasure of public administration. *Public Administration Review*, 76(6), 864–872.
- Lodge, M., & Wegrich, K. (2014). Crowdsourcing and regulatory reviews: A new way of challenging red tape in British government? *Regulation and Governance*, 9, 30–46.
- Margetts, M., & Dunleavy, P. (2013). The second wave of digital-era governance: a quasi-paradigm for government on the Web. *Philosophical Transactions of the Royal Society A*, 371, 1–17.
- McCrummen, S. (2011). D.C. schools to use data from teacher evaluation system in new ways. Washington Post, February 14th, 2011. <http://www.washingtonpost.com/wp-dyn/content/article/2011/02/13/AR2011021302584.html>. Accessed April 25, 2017.
- McDermott, P. (2010). Building open government. *Government Information Quarterly*, 27(4), 401–413.
- McKinsey. (2014). Innovation in local government, open data and information technology’. <http://www.mckinsey.com/industries/public-sector/our-insights/government-designed-for-new-times-number-2>. Accessed April 25, 2017.
- Meier, P. (2013). Can official disaster response Apps Compete with Twitter? iRevolutions, 13 August, 2013. <https://irevolutions.org/2013/08/13/can-official-disaster-response-apps-compete-with-twitter/>. Accessed April 25, 2017.
- Meijer, A., & Bolivar, M. (2015). Governing the smart city: A review of the literature on smart urban governance. *International Review of Administrative Sciences*, 82(2), 392–408.
- Melin, U., Söderström, F., & Axelsson, K. (2016). Managing the development of e-ID in a public e-service context: Challenges and path dependencies from a life-cycle perspective. *Transforming Government People, Process and Policy*, 10(1), 72–98.
- Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *JAMA*, 309(13), 1351–1352.
- Nedlund, A. C., & Garpenby, P. (2014). Puzzling about problems: The ambiguous search for an evidence-based strategy for handling influx of health technology. *Policy Sciences*, 47(4), 367–386.
- Neuman, W. R., Guggenheim, L., Mo Jang, S., & Young Bae, S. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication*, 64, 193–214.
- Nutley, S. M., Walter, I., & Davies, H. T. O. (2007). *Using evidence: How research can inform public services*. Bristol: Policy Press at the University of Bristol.
- OECD. (2014). *Recommendation of the Council on Digital Government Strategies 15 July 2014 - C (2014) 88*. Paris: France.
- Pawson, R. (2006). *Evidence-based policy: A realist perspective*. London: Sage.
- Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51–59.
- Radin, B. (2003). The instruments of intergovernmental management. In P. Guy & J. Pierre (Eds.), *Handbook of public administration* (pp. 606–618). London: Sage.
- Richards, N.M. and J.H. King. (2014). Big Data Ethics. *Wake Forest Law Review*, 49, 393–432.
- Rose, N. (1999). *Powers of Freedom: Reframing Political Thought*. Cambridge: Cambridge University Press.
- Rose, W. R., & Grant, G. G. (2010). Critical issues pertaining to the planning and implementation of E-Government initiatives. *Government Information Quarterly*, 27(1), 26–33.
- Roski, J., Bo-Linn, G. W., & Andrews, T. A. (2014). Creating value in health care through big data: Opportunities and policy implications. *Health Affairs*, 33(7), 1115–1122.
- Sanderson, I. (2006). Complexity, ‘practical rationality’ and evidence-based policy making. *Policy and Politics*, 34(1), 115–132.
- SAS. (2014). *Government workforce in focus: Closing the data and analytics skills gap*. Washington, DC: SAS.
- Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. New York: Doubleday.

- Taylor, M., Rubin, E., & Hounshell, D. (2005). Regulation as the Mother of Innovation: The Case of SO₂ Control. *Law and Policy*, 27(2), 348–378.
- Turque, B. (2012). Creative...motivating and fired. The Washington Post, March 6th, 2012. https://www.washingtonpost.com/local/education/creative-motivating-and-fired/2012/02/04/gIQAwzZpvR_story.html?utm_term=.1781edeed49d. Accessed April 4, 2017.
- Wesselink, A., Colebatch, H., & Pearce, W. (2014). Evidence and policy: Discourses, meanings and practices. *Policy Sciences*, 47(4), 339–344.
- Williamson, B. (2016). Digital education governance: data visualization, predictive analytics, and ‘real-time’ policy instruments. *Journal of Education Policy*, 31(2), 123–141.
- Yiu, C. (2012). The big data opportunity: Making government faster, smarter and more personal. Policy Exchange Report, UK digital government unit, London.