

# Big Data and AI – A transformational shift for government: So, what next for research?

Public Policy and Administration  
0(0) 1–21

© The Author(s) 2018

Reprints and permissions:

[sagepub.co.uk/journalsPermissions.nav](http://sagepub.co.uk/journalsPermissions.nav)

DOI: 10.1177/0952076718780537

[journals.sagepub.com/home/ppa](http://journals.sagepub.com/home/ppa)



**Irina Pencheva**

Government Digital Service, Cabinet Office, UK

**Marc Esteve** 

International Public Management, University College London, UK;  
ESADE Business School, Spain

**Slava Jankin Mikhaylov**

Public Policy and Data Science, University of Essex, UK;  
Essex County Council, UK

## Abstract

Big Data and artificial intelligence will have a profound transformational impact on governments around the world. Thus, it is important for scholars to provide a useful analysis on the topic to public managers and policymakers. This study offers an in-depth review of the Policy and Administration literature on the role of Big Data and advanced analytics in the public sector. It provides an overview of the key themes in the research field, namely the application and benefits of Big Data throughout the policy process, and challenges to its adoption and the resulting implications for the public sector. It is argued that research on the subject is still nascent and more should be done to ensure that the theory adds real value to practitioners. A critical assessment of the strengths and limitations of the existing literature is developed, and a future research agenda to address these gaps and enrich our understanding of the topic is proposed.

## Keywords

Big Data, literature review, policy process

---

## Corresponding author:

Marc Esteve, University College London, 29 Tavistock Square, London WC1H 9QU, UK.

Email: [marc.esteve@ucl.ac.uk](mailto:marc.esteve@ucl.ac.uk)

## Introduction

Big Data is thought to have a global reach and exert a fundamental structural impact throughout society (McNeely and Hahm, 2014). While the use of data in the public sector is not new, the potential and actual use of Big Data applications affects aspects of the theoretical and practical considerations of decision-making in the public sector (Giest, 2017). This is driven not only by the data revolution but also the accompanying development of advanced analytics. From the practitioner perspective, this is summarised in a speech by John Manzoni, Chief Executive of the UK Civil Service and Permanent Secretary of the Cabinet Office on Civil Service Transformation. Civil service is undergoing a transformation with robotics and automation changing the provision of public services. There is a requirement to embrace big data and technology that is reshaping the workforce. The challenge currently is to better use citizen data for improvement of public services, target who needs services more specifically and ‘tailor those services more accurately’ (Manzoni, 2018).

At the most basic level, Big Data is about the volume of information, a variety of the different data sources and types (structured and unstructured), and the velocity – namely the speed of creation, storage and dissemination of data, often in real-time (Einav and Levin, 2014). The three ‘V’s is a term coined to distinguish Big Data from conventional data (Eaton et al., 2012). However, different stakeholders attribute different meanings to the concept (Stough and McBride, 2014). Some see it as ‘a cultural, technological, and scholarly phenomenon’ (Boyd and Crawford, 2012: 663); others as ‘a multidimensional concept embracing technology, decision making, and public policy’ (McNeely and Hahm, 2014: 304). The difficulty of defining such a broad concept has led to several attempts of clarifying its real meaning. As a result, some authors have proposed to include concepts such as veracity, validity, value, and viability – 4Vs (Kimble and Milolidakis, 2015). Although the use of these new concepts has created some controversy, as they do not refer to proportional dimensions of big data, but instead they can refer to all types of data. The very definition of the concept is not the only challenge faced by governments and policymakers; the difficulties ranging from governance and ethical concerns, to structural and organisational resource limitations when dealing with Big Data need to be considered as well (Mergel, 2016; Phillips, 2017; Youtie et al., 2017).

Advanced analytics is related to a more general concept of artificial intelligence (AI) and its underlying technologies. Over the last decade, there have been dramatic advances in core AI technologies like machine learning, natural language processing, virtual agents, and computer vision (Russell and Norvig, 2009). The early promise of AI was largely viewed in terms of providing decision support for public managers (e.g. Hadden, 1986; Hurley and Wallace, 1986; Jahoda, 1986; Masuch and LaPotin, 1989). Latest advances in AI allow computers to learn from past experiences and understand the world through a hierarchy of concepts (Goodfellow et al., 2016) that can lead to automation of tasks (Bailey et al., 2016;

Barth and Arnold, 1999). While many early promises of AI went unfulfilled, recent successful applications represent the third wave of AI that started from around 2006 (Goodfellow et al., 2016). A key contributing factor to increasing maturity of AI technologies and the viability of AI application to public policy and administration is the availability of data that can be used in the computer learning process. At the same time, without the underlying analytical technology, the data revolution can be viewed simply as a shift in the scale of the available data rather than a transformational change. Hence, this study refers to Big Data as a phenomenon where the scale of available data (data revolution, 4Vs, etc.) is integral to machine learning, natural language processing, and other AI technologies. This view is not far removed from the definition of Big Data through the concepts data analytics and data science in Mergel et al. (2016).

Increasing prominence makes it imperative to understand the role of Big Data in the public sector. Despite its importance, research on the topic from a Policy and Administration perspective is still nascent. Maciejewski (2017) provides a valuable summary of the applications and lessons of Big Data in public policy, although this draws largely on practical examples and makes limited mention of the academic understanding of the subject. A notable exception is offered by Mergel et al. (2016), which provide an operational definition of Big Data for public affairs and discuss major challenges for the Public Administration field emanating from Big Data, especially in the field of education. However, there has been little effort to date to consolidate findings, distil the key effects and recommendations for public organisations and, crucially, provide a coherent approach for the future direction of the Public Policy and Administration field. This study aims to address this gap at an opportune moment, given the growing importance of the topic.

## Research design

This study is based on a comprehensive literature review of the role and implications of Big Data for the public sector. It explores the coverage of the subject in leading Policy and Administration journals but also considers the wider field of work, including sources covering social science, management, and information and technology management. This methodological approach was selected as it allows the researcher to (a) identify the central topics and strands of theory; (b) consolidate and critically review our existing knowledge; and (c) provide direction for future research. The impact of Big Data on the public sector is beginning to unfold in the Policy and Administration literature. As a result, an informed, broad and detailed comprehension of the subject is important and timely as it can help to move the discussion forward as to how Big Data can benefit the public sector.

The literature review aims to offer an exhaustive coverage of all articles published in the top 20 journals of the 2018 Google Scholar rankings under the “Public Policy and Administration” subcategory (see Appendix 1 for a list of journals and the tabulation of relevant articles in each journal). These sources were selected

as they provide a credible and, importantly, replicable set of key journals in the discipline. Four additional journals were included to the list for comparability with the Google Scholar rankings from 2017 and 2016 when some of the early research was undertaken (*Journal of Policy Analysis and Management*, *Public Policy and Administration*, *Administrative Science Quarterly*, and *International Journal of Public Sector Management*). The coverage included all articles up to April 2018 that mention any of the following search terms: ‘Big Data’, ‘Data Analytics’, ‘Data Science’, ‘Advanced Analytics’, ‘Machine Learning’, ‘Natural Language Processing’, and ‘Artificial Intelligence’. These terms, in our opinion, capture the essence of the Big Data phenomenon as discussed above and relate to the key descriptors that appear in academic work and, importantly, discussions within the industry and government. An illustrative example here is the reference to AI and data revolution as the first Grand Challenge set out in the UK Government Industrial Strategy (HM Government, 2017). Abstracts (and full articles as needed) were examined to identify studies for this review. In total, 196 articles in primary Public Policy and Administration (henceforth PPA) journals that included a mention of these terms were reviewed. It should be noted that many results were returned that included the search terms but were found not to be related directly to the study topic (e.g. on e-Government, performance, or open data), while some hits contained only in-passing mentions of the terms and did not engage with the subject matter. While not included in the literature review, they were still considered as a background context from which some parallels were drawn that were used for the proposed agenda. Overall, the total number of PPA articles assessed to be substantively relevant and hence included in this study is 57.

To get a holistic understanding of the field, given that the main PPA literature on the subject is still limited, the initial review was further complemented by including articles and books beyond the primary PPA journals. These were selected through an electronic reference search of academic databases – Google Scholar, ProQuest and JSTOR – using a combination of the above key terms and ‘Public Management/administration/government’. Finally, in the third stage, the review included sources that were identified through an analysis of the references in the main Public Administration articles. These second and third stages of the literature review added another 19 substantively relevant articles and 7 books. This approach allowed for insights to be drawn from several fields including political science, organisational theory, management, and wider social sciences. The total number of sources included in this study is 83.

## **Big Data in public administration research**

A major topic in the Public Administration literature, discussed by most scholars, concerns the application and high-level benefits of Big Data for the public sector. There is a consensus in academic research that better use of data would result in several benefits to the public sector, with some authors stating that ‘Big Data has a big future in government’ (Williamson, 2014a: 256), or that it represents a new

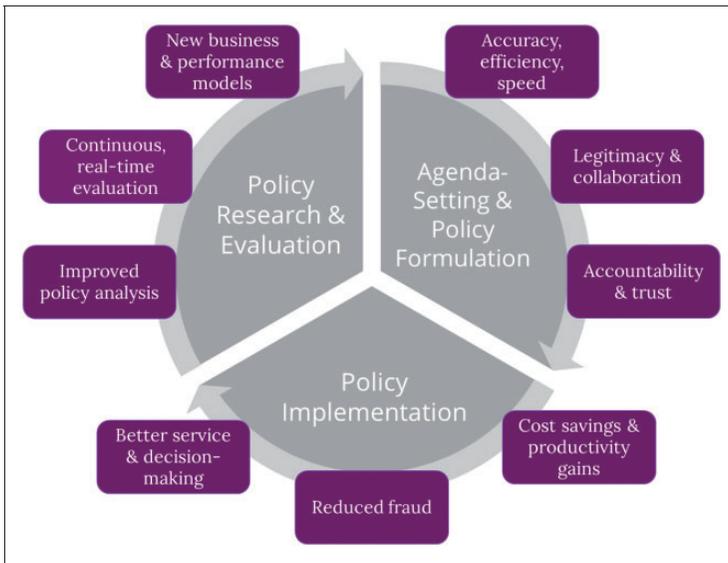
form of competitive differentiation. Most observers confirm that governments are yet to capitalise on the huge potential of this technology (Gamage, 2016). However, there is a growing recognition of the power of data, with public organisations increasingly investing in Big Data solutions and applications (White and Breckenridge, 2014). Analysis of the research field suggests that the key benefits include increased effectiveness, efficiency, and legitimacy of the policy process (Mergel, 2016).

Nevertheless, when it comes to describe the actual applications and advantages, authors diverge considerably in their approach and conclusions. Some academics utilise wide theoretical references to construct their arguments, while others use empirical data such as case studies and interviews to draw assumptions related to a specific area or scenario in Public Administration. The papers cite various examples from country-wide strategies to individual projects and reforms across different functions and levels of government (Gamage, 2016). Some scholars focus on sector-specific applications, most commonly: local and city-level government, taxation and social benefits, and healthcare. This incongruity and diversity in approaches present a significant challenge for observers who wish to fully grasp the applications and benefits of Big Data in government. Given the Public Administration focus of this study, it is out of scope to list all specific developments or feature a detailed discussion for each sector. Instead, this section will identify and analyse the broad themes emerging from the Public Administration and wider literature on the use and advantages of Big Data. It will recall some sector-specific examples to illustrate the theoretical ground.

### *The policy cycle framework*

This study uses the policy cycle framework to reconcile incongruences in the literature and conceptualise the discussion in a systematic and consistent manner. To understand the policy cycle, it is worth considering the definition of public policy. As Birckland (2001) suggests, public policy is a mechanism to contextualise and explain a ‘reaction to an actual, perceived, or anticipated problem’ of societal importance. The policy cycle is a framework to understand the policymaking decision process and methodologically analyse complex strategic activities, actors, and drivers in public policy. In general, the cycle is composed of several inter-related stages, namely agenda-setting, policy formulation, and decision-making, policy implementation, and policy evaluation (Wegrich and Jann, 2007). Each phase is described in greater detail below regarding the specific use of Big Data.

There are several reasons for selecting the ‘policy cycle’ framework as an analytical model to present the findings. First, it offers ‘a basic template [...] to systematise and compare the diverse debates, approaches, and models in the [public policy] field’ (Wegrich and Jann, 2007: 43). As such, it helps to develop a strong policy-driven theoretical perspective and structures the empirical findings in an area of study where Public Administration analysis is still incipient. Second, it



**Figure 1.** Benefits of Big Data in the policy cycle.

detaches the effects of Big Data from any particular type of organisational form or level of government. This means that while there are multiple types of public organisation operating at the local, regional, national or international level, all of these are subject to the design and implementation of public policies. Hence, the use of the policy cycle framework allows us to discuss the effects of Big Data in a wide range of public organisations. Moreover, there are already a few Public Administration scholars who use (parts of) the policy process to describe the application of Big Data (Giest, 2017; Hocht et al., 2016; Schintler and Kulkarni, 2014). In addition, some authors introduce the term ‘policy analytics’ to denote data analytics specifically related to public policy (Daniell et al., 2015). Nonetheless, their conclusions vary significantly, and there is a lack of methodical overview of existing literature. This study aims to build upon their findings and address these gaps. Next, the benefits and applications of Big Data in each phase are described; Figure 1 schematically illustrates these benefits.

### *Agenda-setting and policy formulation*

Agenda-setting is ‘an incremental process of agenda accretion’ whereby multiple public, private and third sector actors engage in a primarily political process to recognise and select an issue for ‘serious consideration of public action’ (O’Jones, 1983 in Princen, 2011: 79; Wegrich and Jann, 2007: 45). It involves framing an issue as a policy problem that requires a public sector solution and setting a policy image of what the required solution should be. Policy formulation includes the

transformation of identified problems and proposals into government programmes. Although closely interrelated and often blurred, the agenda-setting and policy formulation stages are usually treated as discrete periods in the policy study. As the literature on Big Data makes little distinction between the two phases, this study will analyse the applications and benefits across both phases. As discussed below, the benefits of Big Data in agenda-setting and policy formulation are threefold.

First, it may increase the accuracy, efficiency, and speed of the process. Through Big Data and advanced analytics, governments can unlock vast swathes of unstructured data and thereby supplement traditional techniques such as surveys (Bachner and Hill, 2014; Pandey et al., 2017; Zhu et al., 2017). It supports public managers in aggregating and analysing citizens' policy preferences and needs in order to better understand which incentives will work and under what circumstances (Clarke and Margetts, 2014). Moreover, compared to traditional data, Big Data allows for a variety of sources to be researched and analysed to understand the policy landscape (Loftis and Mortensen, 2018; Nowlin, 2016), identify problems, and conceptualise solutions more accurately and in greater detail (Williams, 2014a). In policy formulation, Big Data helps to design a policy that closely matches the preferences (Stritch et al., 2017; Taeihagh, 2017), in addition to developing different scenarios and accurately predicting their possible outcomes (Cook, 2014). This helps to bridge the gap between the original policy objectives and the unintended consequences of implementation – the main challenge in traditional policymaking. The high velocity of data also enables policymakers to quickly (re)act to and incorporate collective information from a variety of sources, including channels outside the formal policy realm such as social media, online consultations and virtual town halls (Hochtl et al., 2016; Mergel, 2017). For example, the Commonwealth Scientific and Industrial Research Organisation uses a social media analysis software tool to quickly alert staff to any changes that may require their attention, such as users who comment on a public post with incorrect statements to block the spread of fake news (Maciejewski, 2017).

Second, Big Data can increase the legitimacy of agenda-setting and policy formulation by enabling citizens and governments to engage in more meaningful dialogue and collaborate in policy design (Desouza and Bhagwatwar, 2012). Some scholars argue that Big Data offers the means for the social and civic empowerment of various non-expert stakeholders. When coupled with other innovative solutions such as civic crowdsourcing and online participatory budgeting, Big Data can enable those with traditionally limited access and little opportunity to express their views to influence the policy process and provide feedback on government decision-making (Mergel, 2016; Taeihagh, 2017). Through Big Data, governments can undertake better assessments of the preferences of citizens. This decreases the reliance on technocratic experts and legitimises the policy process from the start (Schintler and Kulkarni, 2014). Furthermore, regarding collaboration, Big Data helps to identify possible partnerships with actors outside the public sector. For example, improved record checking of private companies at the start of public procurement has enabled 'public agencies [...] to identify the best possible cooperations' (Hochtl et al., 2016: 161).

Third, applying analytics can result in governments being more accountable to citizens and even transform the traditional governance model. Arguably, Big Data has the potential for improving mutual government–citizen understanding and creating a new model of Public Administration for the digital age (Clarke and Margetts, 2014). Goldsmith and Crawford (2014) label this as a new data-smart governance model, and Pisano (2016: 2) refers to a new ‘upside-down thinking [that] creates new rules of the game so that the organisations in our society understand the larger forces they confront and are able to change the way they act to sustain the commons and achieve the social contract’.

Nevertheless, some authors critique the optimism related to the potential of Big Data in the early stages of the policy cycle based on technical, equitable, and wider democratic grounds. First, public policy data, especially at the start of the policy process, is ‘noisy’, with many irregular and heterogeneous properties (Schintler and Kulkarni, 2014). Deriving meaningful insights from it is not only technically difficult but may mislead public managers to pursue policies and allocate resources to chase ghost problems.

Second, from an equity perspective, it is highlighted that there may be limits to the potential for greater social inclusion because the same people who most require empowerment have the least access to technology (Schintler and Kulkarni, 2014). In fact, as Boyd and Crawford (2012: 673) argue, in the current ecosystem it may create or even reinforce new digital divides between the Big Data rich and Big Data poor.

The third is the risk that the cumbersome nature of the democratic process dissuades many citizens from participating in the policy process in the first place. Hence, Big Data’s democratising potential may be overhyped, and its impact may not be so dramatic (Boyd and Crawford, 2012: 14).

### *Policy implementation*

The policy implementation phase refers to the execution or enforcement of a policy by the responsible institutions and organisations that are often, but not always, part of the public sector (Wegrich and Jann, 2007). By far the most exhaustive and mature analysis and coverage of the PPA literature on Big Data refer to this phase. This is not surprising because most of the benefits of this technology accrue on a programmatic level when policy execution takes place. The advantages of Big Data for government in this phase can be broadly grouped in terms of strategy and operations. Big Data has been argued to help the government achieve its policy outcomes (Rogge et al., 2017). For example, data-driven precision governance can enhance hazard preparedness and response efforts (Hondula et al., 2018). In addition, it can influence the strategic decision as to whether to implement a policy through a collaborative venture, such as the co-production project to include citizens in the fight against corruption developed by the US federal government (Bertot et al., 2010).

Despite the importance of Big Data in the strategic decisions faced by public organisations, most research has actually focused on the benefits that can be the

result of placing Big Data at the operational level of policy implementation. In particular, scholars have examined how it can improve efficiency and effectiveness in the delivery of public services (Gamage, 2016). Big Data increases public sector efficiency by delivering savings and boosting productivity (Agostino and Arnaboldi, 2017; Johnes and Ruggiero, 2017). It allows organisations to consolidate information at a lower cost and power up public services that rely heavily on accurate data, such as tax processing. Joseph and Johnson (2013), in their research of the US Department of Veterans Affairs, suggest that the automation and redesign of services through Big Data applications can reduce offline and administrative processing, and optimise current functions. This technology can also help governments to better monitor operational performance and spending to ensure that budget target results are achieved. Techniques like resource allocation modelling and real-time operations optimisation (Daniell et al., 2015) can support public managers in observing and understanding the work of employees. This can serve to expose and eliminate redundancies and inefficiencies in the current methods of policy implementation, such as repeated activities or sub-optimally used resources (Shindelar, 2014). For example, in Singapore, analytics techniques are used to analyse 12 million transactions daily to plan and better operate the transport system (Maciejewski, 2017). Additionally, through Big Data and advanced analytics, important reforms in the government's finance performance could be identified (Mazur, 2015; Sims and Sossei, 2012).

Another main advantage of Big Data is better supervision of the implementation process through the detection of irregularities (Maciejewski, 2017). To that end, using automated algorithms to support human decision-making is suggested to significantly reduce fraud and errors in service processing (Janssen and Kuk, 2016). Examples of these types of applications include Germany's Federal Labour Agency which reduced the case of fraudulent benefits claims by 20%, and the Irish Tax and Customs Authority which used data science techniques to develop predictive models to assist in the better targeting of taxpayers for possible non-compliance/tax evasion, and liquidation (Cleary, 2011).

From an effectiveness point of view, the benefits include enhanced service delivery and improved decision-making (Gamage, 2016). In terms of enhanced delivery, Big Data allows public agencies to better understand and segment users to personalise the interventions. For example, through analytics, some tax agencies have categorised individual and business taxpayers to provide a bespoke offer for each client. Big Data can also be useful in designing new services to address unmet needs. In healthcare, for example, using data from digital health records and tracking policy outcomes can help to develop new interventions for hard-to-reach populations (Blume et al., 2014). In terms of superior insight and decision-making capabilities, Big Data has been used by police departments in the United States to understand and even proactively intervene to prevent crime occurrences at the point of origin (Hochtl et al., 2016).

There are scholars who highlight some limitations of policy implementation. Focusing on instant performance and operational gains and savings could result

in goal displacement or a shift, whereby “public managers and policymakers [...] lose sight of the broader issues the public sector needs to address” (Desouza and Jacob, 2017: 1059). Enabling a pure data culture within the public sector could be antithetic to the concept of delivering public value. Moreover, opening up ‘administrative data and performance information increasingly enables external political actors to peer into and evaluate the administration of public programs’ (Lavertu, 2016: 864). Whilst on the surface this might be a positive development, there is a general risk that these actors are not well-informed or that they are restricted in their grasp of the wider political context and strategic priorities. This could lead to further goal displacement, undue external criticism and policy problems for the public manager.

### *Policy research and evaluation*

The policy evaluation stage aims to assess the intended and unintended consequences of policies and measure these against the original projected purpose and outcomes. As part of that, policy research and analysis is an important activity that contributes to and enables the evaluation process. The theme of policy research features prominently in the PPA literature. Big Data has already transformed the research fields in many natural science disciplines, and proponents call for the same to occur in the PPA arena, as Big Data ‘is ripe for policy analysis’ (Schintler and Kulkarni, 2014: 343). In fact, some authors argue that the benefits and potential of Big Data would be most dramatic at the evaluation phase, so much so that it will completely transform the landscape of social and economic policy and research (Hochtl et al., 2016).

Big Data allows for improved policy analysis (Decker, 2014). First, advanced analytics provide a greater level of granularity, with the ability to simultaneously observe individual and aggregate variables (i.e. regarding neighbourhood, city, local authority, or country, among others). Moreover, the capacity to handle time-series data from multiple, diverse sources supports the holistic measurement of policy outcomes (Jarmin and O’Hara, 2016). These two advantages enable public managers to understand the long-term effects of interventions on citizens in policy areas that have an impact throughout the individual’s life, such as health care and education (Cook, 2014). Additionally, more and more precise information for both quantifiable (e.g. economic costs) and non-quantifiable (e.g. ethical or moral concerns) factors can be handled. This could be applied in sophisticated cost-benefit analyses that adequately represent ‘the monetised long-term benefits [of a project or policy] for individuals and society’ (Cook, 2014: 544).

Although it is usually visualised as the end of the traditional policy cycle, often some evaluation occurs during each stage of the process. This is further reinforced by Big Data which ‘open[s] permanent possibilities of reiteration, reassessment, and consideration’ of policy, what is called the e-policy cycle by Hochtl et al. (2016). Big Data allows for a rapid, even real-time evaluation process because ‘the very

execution of new policies will almost immediately produce new data, which can then be used to evaluate the effectiveness of these policies and enhance future implementation' (Hochtl et al., 2016: 151). In fact, the focus on continuous research and evaluation has resulted in policies and programmes featuring a focus on the data that will be produced and how it should be analysed (Pirog, 2014).

Finally, Big Data also enables experimentation with new business models and organisational performance techniques. As Arinder (2016: 394) suggests, public sector institutional culture can change and become more predictable and transparent through 'data and time-conscious evaluative frameworks that emphasise evidence-based decision making and longitudinal cost-benefit analytics at critical policy-making junctures'. Overall, Big Data could allow public managers to significantly shorten the feedback loop for employee performance and try out new business models to improve operational performance.

As with the other stages of the policy process, Big Data could also have a downside for policy research and evaluation. The focus on data-driven policy interventions could lead to the dominance of data over substantive theory in the policy process (Cook, 2014). As White and Breckenridge (2014) warn, Big Data is not a panacea for all Public Administration problems. Solely focusing on evaluative processes introduces the risk of addressing immediate problems but missing the underlying causes. In addition, 'data for policy decisions often do not connect with data for policy implementation – and vice versa', making it difficult to complete an overall evaluation of the policy (Kettl, 2016: 4). To exemplify, using analytics, the New York Office of Policy and Strategic planning uncovered a 'curious correlation' between a building type and illegal conversions, and swiftly addressed the issue. However, as Desouza and Jacob (2017) underscore, whilst the issue was eradicated, the underlying social problem (of poverty and social exclusion) is likely to remain.

There is a huge drive for governments to collect and extract as much value as possible from all kinds of data (Boyd and Crawford, 2012). However, in substance, if the quality of this data is poor Big Data becomes obsolete, what many sceptics refer to as garbage in, garbage out. Simultaneously, data cleaning can be costly and time consuming; thus, reiterating the need to only use Big Data for a worthy purpose. Data quality becomes a crucial component of good policy research and evaluation. Therefore, Public Administration should not rely solely on numbers since these are far more fallible than we believe (Mayer-Schonberger and Cuckier, 2013).

## **Challenges for adoption**

Public organisations generate huge amounts of data; yet they are often unlikely to use it to gather valuable insights or transform services. Another key topic in the PPA literature tackles the reasons for this mismatch between the simultaneous overproduction and underconsumption of data in government (Joseph and

Johnson, 2013). It recognises that realising the potential of Big Data is not problem-free and analyses the barriers to adoption in the public sector. These challenges range from technical and application risks to moral objections. Alongside the barriers, there 'are serious and wide-ranging implications for the operationalisation of Big Data' in the public sector (Boyd and Crawford, 2012: 675).

To structure the issues and implications in a way that supports easier understanding for the public sector context, we opted to detach them from the policy cycle stages. This is arguably because the challenges described below will influence several stages of this framework. So, instead of using the policy cycle framework to describe the challenges faced by public organisations when implementing Big Data-related initiatives, this study adopts three levels of analysis, namely: system, organisational, and individual. These tiers have been used by scholars to analyse a range of subjects related to traditional PPA such as performance (Ashworth et al., 2013), and collaborations (Esteve et al., 2012), in addition to exploring emerging themes like open data (Wirtz et al., 2015). More recently, Allard et al. (2018) used the tiered approach to analyse challenges in administrative data use within US government agencies. Thus, applying the same cognitive structures and labels allows the reader to draw comparative conclusions between Big Data and other similar topics of importance to PPA.

### *System level*

In the context of this study, system-level barriers are understood to be those that arise from the 'networked' nature of Big Data itself, impacting the technology's adoption (Desouza and Jacob, 2017). They are challenges not specific to an individual public sector organisation, but rather those that hinder development across government more generally. Based on the literature, several interrelated challenges can be identified at this level regarding privacy and security, data governance and ethics. These issues are discussed at length in the literature but are often also referenced in passing across most papers as a sober reminder of the potential limitations.

Big Data presents some challenges applicable to both public and private organisations. Fundamentally, it elevates the concern about personal and organisational privacy. Analysed and collected data often contains individually identifiable information – i.e. Big Meta Data – that, even when anonymised, could be attributed back to users (Stough and McBride, 2014). Without robust data governance principles that establish the purpose of collected data and the principles for its (re)use, it is possible to connect seemingly unrelated data-points to grasp significant information and even the identity of an individual without their consent. Consent itself becomes a muddled concept in the Big Data age as 'it is not easy to opt out from a dataset and the act of opting out might identify a person' (Peled, 2013: 8). To that end, many observers highlight the challenge of establishing legislation that would address these complex eventualities (Combe, 2015). Intricately related is the question of ethical use (Hondula et al., 2018). As Boyd and Crawford (2012) note, just because data is accessible does not make its use ethical. The problem is particularly

important for sectors such as healthcare, where data from fitness tracker devices could be used by insurers to reduce information asymmetries but could also negatively impact the individual's insurance options (if insurers are aware of their lifestyle) (Mergel, 2016). Finally, in the age of Big Data, more organisational and personal data is at stake, which intensifies the security risks.

Several factors within the public sector intensify these systemic issues. First, due to their mandates, public agencies collect vast swathes of sensitive information on citizens, ranging from healthcare records to social benefits. These, if used together, could elucidate an almost complete picture of the individual's life, significantly undermining privacy. Second, organisations are permitted to collect only the data needed to fulfil their missions. In government, where mandates are often unclear, unstable, and subject to multiple interpretations, having clear purpose limitation rules becomes even more problematic. Third, there are wide-ranging, often dystopian speculations around the ethical use of Big Data in government. If sophisticated predictive models can be used to forecast crimes before they happen, would it be just or equitable to use this information in probabilistic policymaking to pre-emptively punish potential offenders (Clarke and Margetts, 2014). Some sceptics argue that Big Data in the public sector could lead to 'mass surveillance' and 'an Orwellian state monitoring of citizens' (Williamson, 2014b: 255). For example, real-time censoring techniques and analytics on policy preferences have been used in China to identify and warn policymakers of potential salient issues and political unrest (King et al., 2013). Finally, regarding security, public organisations have become a regular target for cybercriminals due to the valuable data that is transacted, and in the Big Data world this threat would be even greater (McNeely and Hahm, 2014).

Despite the attention directed towards this topic, significant questions remain around addressing the system-level barriers. Overall, the Public Policy and Administration literature offers little concrete or operationally feasible recommendations, and these seem to be more of a 'wish-list' that varies significantly across publications. A notable exception is the discussion in Chen and Lee (2018) that brings together perspectives from the literature on e-government, collaborative governance, and network management to propose best-practice solutions for collaborative data networks regarding information and decision-support. To tackle the issues around privacy and the ethical use of Big Data, it has been suggested that policies and legislation fit for the Big Data age should be designed (Desouza and Jacob, 2017). A new legal framework should be introduced to govern the ever-expanding variety of statistical sources and data services at government disposal (Washington, 2014). A 'data deluge' could significantly increase the opportunities for abuse and misuse, so ethical use becomes imperative (Kuiler, 2014). Pyrozhenko (2017: 1509) also calls for 'the institution of public administration [to] play an active role in managing' the development of the trend and setting purpose limitation rules. To overcome data governance issues, Lane (2016) suggests the establishment of a quadruple helix model, made up of state and city agencies, universities, private data providers, and federal agencies with each

actor having important responsibilities in the data management and exchange process. On a technical level, it is recommended to invest in research on identifying a technology for stripping personally identifiable information (Stough and McBride, 2014). Overall, as Schintler and Kulkarni (2014) note, designing and implementing public policies that tackle the negative implications of Big Data would be a very complex task, but one worth pursuing.

### *Organisational level*

The second type of barriers that impact the adoption of Big Data in government is organisational-level constraints – namely around collaboration, resources and skills (Allard et al., 2018). Data creates unique challenges for collaboration in the public sector for several reasons. First, the lack of policy and regulatory frameworks to guide and promote collaboration. In fact, poor data governance across organisational boundaries is viewed by public managers as a significant barrier to adoption (Desouza and Jacob, 2017). From an operational perspective, the relatively siloed approach within which many public sector organisations operate causes a range of issues – from the technical interoperability of IT systems to the lack of comparable data parameters (Hochtl et al., 2016). It is also suggested that ‘coordination costs associated with data-sharing may be greater than other types of collaborative endeavours in the public sector’ (Hochtl et al., 2016). Peled (2013) argues that turf wars might be another significant hindrance to data sharing across bureaucratic organisations. He suggests that Big Data is used by agencies as a ‘weapon to fight over funds, influence, and autonomy’, and references the arguments of the American Government Accountability Office that ‘the continuous refusal of agencies to share their Big Data is one reason why US exports are not as competitive as they can be’ (Peled, 2013: 9).

More generally, organisational culture and inertia in some public sector organisations are also seen as limiting the adoption of new strategies to derive value from data (Joseph and Johnson, 2013). Simultaneously, election cycles and the resultant changes to the political authorising environment could also impact the momentum and pace of Big Data transformation. From an operational standpoint, there is a relative lack of resources and skills within public organisations to effectively implement Big Data solutions. Arguably, government bodies have a ‘dubious record on the guardianship of large-scale datasets, the management of contract relationships and large technology-based projects, and capacity to innovate with newer media and technologies in comparison with firms, third sector organisations, and citizens themselves’ (Clarke and Margetts, 2014: 409). Hence, they also lag behind in Big Data, although this situation is changing with technological advancements in developing data-driven solutions for policymakers (Isett and Hicks, 2018). In a survey of public managers about the barriers to adopting data analytics, Sims and Sossei (2012) found that the key challenge identified was inadequate budget resources, closely followed by lack of appropriate staff. Unleashing Big Data and using it effectively also requires strong expertise and a deep knowledge

of technology and organisational processes (McNeely and Hahm, 2014). Yet, the benefits of data-driven decision-making are still not fully recognised in many public organisations, making investment decisions for Big Data more complicated (Desouza and Jacob, 2017).

The implications identified at organisational level relate to the development of functional capabilities and wider organisational culture. Operationally, the suggestions for public organisations relate closely to the identified barriers. A basic recommendation is to not only ‘invest in big data skills [but also] institutional[ise] capacity building or incorporate such skills into existing standard operating procedures’ (Mergel, 2016: 234). It is also necessary to advance the technical infrastructure used by the government for Big Data management, for example through investment in analytics and warehouse optimisation (Joseph and Johnson, 2013). It is important for ‘organisational capacity building [...] to go beyond the IT department and include change managers as well as data scientists in the redesign of processes and systems, so that such systems reflect the changing needs of stakeholders’ (Joseph and Johnson, 2013: 238). This requires a more fundamental transformation towards an organisational culture that entrenches Big Data at the heart of processes. In this line, Arinder (2016) calls for evidence-sensitive policymaking, which puts data as the foundation for change and decisions.

### *Individual level*

Finally, a few barriers at the individual level are noted in the literature, but relatively little attention is paid to these. The main challenges in this area are identified for those public managers and policymakers with decision-making power. Particularly, the attitude of public managers towards risk is an important factor, whereby more risk-averse individuals might be less likely to effectively adopt and utilise Big Data (Wirtz et al., 2015). Furthermore, Maciejewski (2017) argues that individual data owners might resist calls to collaborate for personal reasons. Not specific to Big Data, Weitzman et al. (2006) also recognise inconsistent leadership and the absence of political will as challenges to using data to improve policymaking.

Some scholars argue that public managers must fully embrace and understand the potential of Big Data for their work (Shindelar, 2014). Mergel (2016) also suggests that managers should consider and learn more about all dimensions of Big Data, outlining five key elements – ethics, technology, process, organisational and institutional change, and analytics. Individuals should invest in gaining the knowledge in order to use and manage it appropriately. On a cognitive level, adopting a data-driven mindset would be challenging and profound; ‘most people base their decisions on a combination of facts and reflection, plus a heavy dose of guesswork’ (Mergel, 2016: 144). Fundamentally, public managers should adopt a big data mindset, constantly searching for opportunities to unlock new forms of value from data (Mayer-Schoenberger and Cukier, 2013).

## Conclusion

Scholars have devoted significant effort to listing and understanding the practical applications and gains from Big Data throughout the policy process. However, limitations were identified in the development of new theories to explain the use of Big Data in the public sector and in the updating of existing models. This study identified the theories of policy analytics and the e-policy cycle as novel to Big Data. Often, researchers have either referred *en passe* to existing theories such as these relating to e-government or exclusively focused on the applied discussion without any theoretical grounding. It is recognised that constructing theories is not necessary to provide valuable analysis and should only be undertaken with a clear purpose. Nonetheless, in the underdeveloped area of study that is Big Data in the public sector, it would be beneficial to have advanced models (even if they are built on existing ones) to better understand and assess the field. Furthermore, having a solid theoretical foundation is essential to draw parallels and lessons across different contributions, and provide conclusions at cross-governmental level.

This study is not without limitations. One main constraint is that the use of the policy cycle framework to articulate the discussed arguments fails to tap into some of the more context-specific findings disclosed by Big Data research. In this sense, further research should consider all possible contingencies when describing the possible effects of Big Data-related initiatives on specific public policies or public organisations.

Big Data has profound potential to support public sector transformation, and it is thus paramount to continue advancing and deepening our understanding of the subject. This study has outlined the need for PPA scholars to take up this task and support public managers in their quest for reform. However, governments should also support the efforts for knowledge creation and analysis by opening up their data further, collaborating with – and actively seeking inputs from – researchers to understand how Big Data can be utilised in the public sector. Ultimately, the supporting field of academic thought will only be as strong as the public administration practice allows it to be.

## 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 authors gratefully acknowledge the financial support of the Spanish Ministerio de Economía y Competitividad. Ref: COLLABdet CSO2016-80823-P.

## ORCID iD

Marc Esteve  <http://orcid.org/0000-0002-9732-8082>

## References

- Agostino D and Arnaboldi M (2017) Social media data used in the measurement of public services effectiveness: Empirical evidence from Twitter in higher education institutions. *Public Policy and Administration* 32: 296–322.
- Allard S, Wiegand E, Schlecht C, et al. (2018) State agencies' use of administrative data for improved practice: Needs, challenges, and opportunities. *Public Administration Review* 78(2): 240–250.
- Arinder M (2016) Bridging the divide between evidence and policy in public sector decision making: A practitioner's perspective. *Public Administration Review* 76(3): 394–398.
- Ashworth R, Ferlie E, Hammerschmid G, et al. (2013) Theorizing contemporary public management: International and comparative perspectives. *British Journal of Management* 24(S1): 1–17.
- Bachner J and Hill K (2014) Advances in public opinion and policy attitudes research. *Policy Studies Journal* 42(S1): 51–70.
- Bailey S, Anttiroiko AV and Valkama P (2016) Application of Baumol's Cost Disease to Public Sector Services: Conceptual, theoretical and empirical falsities. *Public Management Review* 18(1): 91–109.
- Barth T and Arnold E (1999) Artificial intelligence and administrative discretion. *The American Review of Public Administration* 29(4): 332–351.
- Bertot JC, Jaeger PT and Grimes JM (2010) Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly* 27(3): 264–271.
- Birckland T (2001) *An Introduction to the Policy Process: Theories, Concepts, and Models of Public Policy Making*. New York, NY: ME Shape.
- Blume G, Scott T and Pirog M (2014) Empirical innovations in policy analysis. *Policy Studies Journal* 42(S1): 33–50.
- Boyd D and Crawford K (2012) Critical questions for big data. *Information, Communication & Society* 15(5): 662–679.
- Chen Y-C and Lee J (2018) Collaborative data networks for public service: governance, management, and performance. *Public Management Review* 20(5): 672–690.
- Clarke A and Margetts H (2014) Governments and citizens getting to know each other? Open, closed, and big data in public management reform. *Policy & Internet* 6(4): 393–417.
- Cleary D (2011) Predictive analytics in the public sector: Using data mining to assist better target selection for audit. *Electronic Journal of e-Government* 9(2): 132–140.
- Combe C (2015) Privacy, big data, and the public good: Frameworks for engagement. *Local Government Studies* 41(1): 181–182.
- Cook T (2014) "Big data" in research on social policy. *Journal of Policy Analysis and Management* 33(2): 544–547.
- Crozier M (2015) Governing codes: Information dynamics and contemporary coordination challenges. *Administration & Society* 47(2): 151–170.
- Daniell K, Morton A and Rios ID (2015) Policy analysis and policy analytics. *Annals of Operations Research* 236(1): 1–13.
- Decker P (2014) Presidential address: False choices, policy framing, and the promise of "big data." *Journal of Policy Analysis and Management* 33(2): 252–262.
- Desouza K and Jacob B (2017) Big data in the public sector: Lessons for practitioners and scholars. *Administration & Society* 49(7): 1043–1064.
- Desouza K and Bhagwatwar A (2012) Leveraging technologies in public agencies: The case of the U.S. Census Bureau and the 2010 Census. *Public Administration Review* 72(4): 605–614.

- Eaton C, DeRoos D, Deutsch T, et al. (2012) *Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data*. New York: The McGraw-Hill Companies.
- Einav L and Levin J (2014) The data revolution and economic analysis. In: Lerner J and Stern S (eds) *Innovation Policy and the Economy*. vol.14. Chicago, IL: University of Chicago Press, pp.1–24.
- Esteve M, Boyne G, Sierra V, et al. (2012) Organizational collaboration in the public sector: Do chief executives make a difference? *Journal of Public Administration Research and Theory* 23(4): 927–952.
- Gamage P (2016) New development: Leveraging ‘big data’ analytics in the public sector. *Public Money & Management* 36(5): 385–390.
- Giest S (2017) Big data for policymaking: Fad or fasttrack? *Policy Sciences* 50(3): 367–382.
- Goldsmith S and Crawford S (2014) *The Responsive City: Engaging Communities Through Data-Smart Governance*. San Francisco, CA: JosseyBass.
- Goodfellow I, Bengio Y and Courville A (2016) *Deep Learning*. Cambridge, MA: MIT Press.
- Hadden S (1986) Intelligent advisory systems for managing and disseminating information. *Public Administration Review* 46: 572–578.
- Hondula D, Kuras E, Longo J, et al. (2018) Toward precision governance: Infusing data into public management of environmental hazards. *Public Management Review* 20(5): 746–765.
- HM Government. (2017) *Industrial Strategy Building a Britain fit for the future*. White Paper.
- Höchtel J, Parycek P and Schöllhammer R (2016) Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce* 26(1–2): 147–169.
- Hurley M and Wallace W (1986) Expert systems as decision aids for public managers: An assessment of the technology and prototyping as a design strategy. *Public Administration Review* 46: 563–571.
- Isett K and Hicks D (2018) Providing public servants what they need: Revealing the “unseen” through data visualization. *Public Administration Review* 78(3): 479–485.
- Jahoda M (1986) Artificial intelligence: An outsider’s perspective. *Science and Public Policy* 13(6): 333–340.
- Jarmin R and O’Hara A (2016) Big data and the transformation of public policy analysis. *Journal of Political Analysis and Management* 35(3): 715–721.
- Janssen M and Kuk G (2016) The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly* 33(3): 371–377.
- Johnes G and Ruggiero J (2017) Revenue efficiency in higher education institutions under imperfect competition. *Public Policy and Administration* 32(4): 282–295.
- Joseph R and Johnson N (2013) Big data and transformational government. *IT Pro*. pp.43–48. Available at: <http://computer.org/ITPro> (accessed 28 August 2016).
- Kettl D (2016) Making data speak: Lessons for using numbers for solving public policy puzzles. *Governance* 29(4): 573–579.
- Kimble C and Milolidakis G (2015) Big Data and business intelligence: Debunking the myths. *Global Business and Organizational Excellence* 35(1): 23–34.

- King G, Pan J and Roberts M (2013) How censorship in China allows government criticism but silences collective expression. *American Political Science Review* 107(2): 326–343.
- Kuiler E (2014) From big data to knowledge: An ontological approach to big data analytics. *Review of Policy Research* 31(4): 311–318.
- Lane J (2016) Big data for public policy: The quadruple helix. *Journal of Policy Analysis and Management* 35(3): 708–715.
- Lavertu S (2016) We all need help: “Big data” and the mismeasure of public administration. *Public Administration Review* 76(6): 864–872.
- Loftis M and Mortensen P (2018) Collaborating with the machines: A hybrid method for classifying policy documents. *The Policy Studies Journal*. DOI: 10.1111/psj.12245.
- Maciejewski M (2017) To do more, better, faster and more cheaply: Using big data in public administration. *International Review of Administrative Sciences* 83(1): 120–135.
- Manzoni J (2018) Civil service transformation speech. London, 24 January 2018. Available at: [www.gov.uk/government/speeches/civil-service-transformation-speech](http://www.gov.uk/government/speeches/civil-service-transformation-speech) (accessed 10 April 2018).
- Masuch M and LaPotin P (1989) Beyond garbage cans: An AI model of organizational choice. *Administrative Science Quarterly* 34(1): 38–67.
- Mayer-Schoenberger V and Cukier K (2013) *Big Data – A Revolution that will Transform How We Live, Work, and Think*. Boston: Eamon Dolan Book.
- Mazur E (2015) Data Analytics – A Tool for Building Trust in Government. *Journal of Government Financial Management* 64(2): 35–39.
- McNeely C and Hahm J (2014) The big (data) bang: Policy, prospects, and challenges. *Review of Policy Research* 31(4): 304–310.
- Mergel I (2016) Big data in public affairs education. *Journal of Public Affairs Education* 22(6): S231–S248.
- Mergel I (2017) Building holistic evidence for social media impact. *Public Administration Review* 77(4): 489–495.
- Mergel I, Rethemeyer K and Isett K (2016) Big data in public affairs. *Public Administration Review* 76(6): 928–937.
- Nowlin M (2016) Modeling issue definitions using quantitative text analysis. *The Policy Studies Journal* 44(3): 309–331.
- Pandey S, Pandey SK and Miller L (2017) Measuring innovativeness of public organizations: Using natural language processing techniques in computer-aided textual analysis. *International Public Management Journal* 20(1): 78–107.
- Peled A (2013) *The Politics of Big Data – A Three-Level Analysis*. Paper presented at the *European consortium of political research (ECPR) general conference*, Bordeaux, France, 4–7 September.
- Phillips F (2017) A perspective on ‘big data’. *Science and Public Policy* 44(5): 730–737.
- Pirog M (2014) Data will drive innovation in public policy and management research in the next decade. *Journal of Policy Analysis and Management* 33(2): 537–543.
- Pisano M (2016) How research can drive policy: Econometrics and the future of California’s infrastructure. *Public Administration Review* 76(4): 538–539.
- Princen S (2011) Agenda-setting strategies in EU policy processes. *Journal of European Public Policy* 18(7): 927–943.

- Pyrozhenko V (2017) Open government: Missing questions. *Administration & Society* 49(10): 1494–1515.
- Rogge N, Agasisti T and De Witte K (2017) Big data and the measurement of public organizations' performance and efficiency: The state-of-the-art. *Public Policy and Administration* 32(4): 263–281.
- Russell S and Norvig P (2009) *Artificial Intelligence: A Modern Approach*, 3rd ed. Harlow, Essex: Prentice Hall.
- Shindelar S (2014) Big data and the government agency. *Public Manager* 43: 52–56.
- Schintler L and Kulkarni R (2014) Big data for policy analysis: The good, the bad, and the ugly. *Review of Policy Research* 31(4): 343–348.
- Sims H and Sossei S (2012) Federal government primed for data analytics. *The Journal of Government Financial Management* 61: 34–37.
- Stough R and McBride D (2014) Big data and U.S. public policy. *Review of Policy Research* 31(4): 339–342.
- Stritch J, Pedersen M and Taggart G (2017) the opportunities and limitations of using Mechanical Turk (MTURK) in public administration and management scholarship. *International Public Management Journal* 20(3): 489–511.
- Taeihagh A (2017) Crowdsourcing: A new tool for policy-making? *Policy Sciences* 50(4): 629–647.
- Washington A (2014) Government information policy in the era of big data. *Review of Policy Research* 31(4): 319–325.
- Wegrich K and Jann W (2006) Chapter 4: Theories of the policy cycle. In: Fischer F, Miller GJ and Sidney MJ (eds) *Handbook of Public Policy Analysis. Theory, Politics, and Methods*. Boca Raton: CRC Press, pp. 43–62.
- Weitzman B, Silver D and Brazill C (2006) Efforts to improve public policy and programs through data practice: Experiences in 15 distressed American cities. *Public Administration Review* 66: 386–399.
- White P and Breckenridge R (2014) Trade-offs, limitations, and promises of big data in social science research. *Review of Policy Research* 31(4): 331–338.
- Williamson A (2014a) Big data and the implications for government. *Legal Information Management* 14(4): 253–257.
- Williamson B (2014b) Knowing public services: Cross-sector intermediaries and algorithmic governance in public sector reform. *Public Policy and Administration* 29(4): 292–312.
- Wirtz B, Piehler R, Thomas M, et al. (2015) Resistance of public personnel to open government: A cognitive theory view of implementation barriers towards open government data. *Public Management Review* 18(9): 1335–1364.
- Youtie J, Porter A and Huang Y (2017) Early social science research about big data. *Science and Public Policy* 44(1): 65–74.
- Zhu J, Huang H and Zhang D (2017) “Big tigers, big data”: Learning social reactions to China's anticorruption campaign through online feedback. *Public Administration Review*.

## Appendix I List of Public Policy and Administration journals

The primary Public Policy and Administration journals were established based on the top 20 ranking of Public Administration Journals by Google Scholar, as of 8 April 2018. Table 1 provides a list of all the primary journals considered and the number of articles addressing one of the key search terms.

**Table 1.** List of primary Public Policy and Administration journals.

Journal name	Number of articles
Public Administration Review	14
Review of Policy Research	7
Journal of Policy Analysis and Management	5
Public Policy and Administration	4
Policy Studies Journal	4
Administration & Society	4
Public Management Review	4
Science and Public Policy	3
International Public Management Journal	3
The American Review of Public Administration	2
Policy Sciences	2
Journal of Public Administration Research and Theory	1
Governance	1
International Review of Administrative Sciences	1
Local Government Studies	1
Administrative Science Quarterly	1
Public Administration	0
Environment and Planning C: Government and Policy	0
Social Policy & Administration	0
Policy & Politics	0
Review of Public Personnel Administration	0
Public Administration and Development	0
Publius: The Journal of Federalism	0
International Journal of Public Sector Management	0
<b>Total</b>	<b>57</b>

Note: The journals are ordered by the number of articles addressing at least one of the key terms.