

# Big data and the measurement of public organizations' performance and efficiency: The state-of-the-art

Citation for published version (APA):

Rogge, N., Agasisti, T., & 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. <https://doi.org/10.1177/0952076716687355>

## Document status and date:

Published: 01/10/2017

## DOI:

[10.1177/0952076716687355](https://doi.org/10.1177/0952076716687355)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

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

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

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Download date: 25 Apr. 2024

# Big data and the measurement of public organizations' performance and efficiency: The state-of-the-art

**Nicky Rogge**

KU Leuven, Belgium

**Tommaso Agasisti**

School of Management, Politecnico di Milano, Italy

**Kristof De Witte**

Leuven Economics of Education Research, KU Leuven Top Institute for Evidence Based Education Research, University of Maastricht, Netherlands

Public Policy and Administration

2017, Vol. 32(4) 263–281

© The Author(s) 2017

Reprints and permissions:

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

DOI: 10.1177/0952076716687355

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



## Abstract

The increasing availability of statistical data raises opportunities for 'big' data and learning analytics. Here, we review the academic literature and research relating to the use of big data analytics in the public sector, and its contribution to public organizations' performance and efficiency. We outline the advantages as well as the limitations of using big data in public sector organizations and identify research gaps in recent studies and interesting areas for future research.

## Keywords

Big data, efficiency analysis, learning analytics, performance analysis, public sector performance

---

## Corresponding author:

Tommaso Agasisti, Via Lambruschini 4b, Milan 20156, Italy.

Email: [tommaso.agasisti@polimi.it](mailto:tommaso.agasisti@polimi.it)

## Introduction<sup>1</sup>

The persistent and global financial crisis, and attendant strategies for fiscal consolidation have accentuated the problem of maximizing public administration (PA) efficiency. In many areas of public activities – especially in the social welfare domain (e.g. health care, education, elderly care) – securing increased efficiency and productivity ensure resources are focused on improving the quality of activities and outputs, guaranteeing the satisfaction of citizens who receive the services, and assuring effectiveness and equity in the public sector as a whole.

The literature addressing the empirical measurement of public sector efficiency employs a wide range of techniques and focuses on various units of analysis: local governments (Asatryan and De Witte, 2015; De Borger and Kerstens, 1996; De Witte and Geys, 2011; De Witte and Moesen, 2010; Revelli and Tovmo, 2007), public hospitals (Hollingsworth, 2008), public agencies (Neshkova and Guo, 2012), public transportation services (Pina and Torres, 2001; Sampaio et al., 2008), and, education (Cherchye et al., 2010; De Witte and Lopez-Torres, 2017).<sup>2,3</sup>

The application of empirical models for assessing the efficiency of public entities can also open the door to the study of its determinants, and consequently have interesting implications for policy, administration and management of the public services. In this context, for instance, it can be tested whether particular managerial tools, different roles for the regulations, or stimulating policies and interventions (for instance, favoring competition, or facilitating strategic management processes) have a positive or negative impact on the efficiency of public spending (e.g. see De Witte and Geys, 2011, for a discussion on the preferences of voters in local government efficiency).

The new opportunities offered by big data can help the efficiency analysis of public entities make a further step.<sup>4</sup> More specifically, nowadays, administrative datasets are ‘big’ in the sense that the individual organizations, in many sectors, periodically produce very detailed questionnaires and databases that include structural or ‘hard’ information and soft data about managerial practices, quality of outputs and inputs, etc. In addition, a huge amount of information is released by individual public organizations, and can be collected as open data. Finally, the diffusion of e-government practices implies the production of huge amount of data through, for instance, the social networks, open data platforms, and public agencies websites.

Yet, the way in which governments, policy makers, and public organizations use big data and learning analytics is still an under-investigated topic, as is the possibility of using big data for enriching efficiency analyses and performance measurements. This review and the symposium of papers in this journal issue, examine both the use of big data for efficiency purposes and the effectiveness of public and welfare services, and for measuring performance in a broader sense.

We are guided by four research foci:

1. From a theoretical perspective, *what are the advantages of using big data in the understanding of public sector organizations?* Can new available datasets help the

- organizations in designing new services, better evaluating their activities, meeting new and more articulated needs, improving the efficiency of operations?
2. *How do public administrations use big data for their internal performance management procedures?* Is big data employed for comparing outputs, practices (processes) and resources invested, with an explicit aim of benchmarking with similar organizations? How can using big data improve such benchmarking?
  3. *Can big data help the development of new indicators for outputs and inputs, thus allowing innovative efficiency analyses, which can be used to challenge the existing evidence about the efficiency of public administrations?* How do these new studies change the implications that derive from existing literature in the field?
  4. *Are public policy-makers using big data for designing policies and/or adjusting them, for example following the judgments of citizens that can be processed via adequate analytics?*

The main objective of this review is to give an overview of the academic literature and research related to the theme of big data analytics for public organizations' performance and efficiency measurement, with specific attention to our four themes. To conduct this research, we focus on the most recent studies in leading journals that publish on the relationships between information technology, (public) policy making, PA and government (predominantly the journals *Public Policy and Administration* and *Government Information Quarterly*). We also look at recent research report by important political and consultancy institutions (European Commission, McKinsey Global Institute) and recent books on the topic. For the big data applications, we take a broader look at the literature. It is important to note that the overview is not intended to be comprehensive.

This article is the first within a symposium in 'Public Policy and Administration' on 'Big data analytics and its use in the measurement of public organizations' performance and efficiency'. The next two papers of the symposium deal with innovative applications. The paper by Johnes and Ruggiero (2017) focuses on revenue efficiency, in particular ascertaining the extent to which, given output prices, producers choose the revenue maximizing vector of outputs. They evaluate efficiencies for English institutions of higher education for the academic year 2012–13 and find considerable variation across institutions in revenue efficiency. The relaxation of the price-taking assumption leads to relatively small changes, in either direction, to the estimated revenue efficiency scores. A number of issues surrounding the modeling process are raised and discussed, including the determination of the demand function for each type of output and the selection of inputs and outputs to be used in the model.

The third paper of the symposium is by Agostino and Arnaboldi (2017). They show how social media data represent a potential powerful tool in the hands of public authorities to support the evaluation of public service performance. By relying on an action research project in the higher education field, this study explores how social media data can contribute to measure service effectiveness by focusing specifically on Twitter in the higher education field. The aim of the

paper is to develop a set of measures, derived from Twitter data, to quantify the effectiveness of higher education services. This investigation supports a broader discussion about the extent to which social media data can contribute to performance measurement in the public sector.

The article at hand has five main goals. First, it provides readers with a general introduction to the topic area: in particular, it aims to give a clear understanding of the most prescient insights, opinions, results, and big data applications for the public sector that have been described in the literature. Second, a special focus in the review will be on the advantages as well as the limitations of using big data in public sector organizations. Third, the review briefly describes what past studies have written about the use of big data by governments and PAs for internal performance management. In this regard, a particularly interesting research question is whether managers and heads of department have used big data to develop new or better versions of performance indicators (inputs, outputs, and/or outcomes), and/or have used information generated through big data for improving policy making and managerial practices. Fourth, the review considers the potential benefits and applications of big data for commerce and industry (in a context of providing services to the government or not) – this insight is helpful in detecting factors that are growingly important also for the public sector. Finally, the review identifies research gaps in recent studies and fruitful areas for future research, with the aim of setting a tentative agenda for interested scholars.

## **Describing big data**

In general, big data refers to huge volumes of (digital) data that are collected from large variety of sources that are too large, raw, or unstructured for analysis through conventional database techniques (Kim et al., 2014: 78). A common framework that is used to describe big data is the ‘3-V’ framework with the three dimensions ‘Volume’, ‘Variety’, and ‘Velocity’ (Brynjolfsson and McAfee, 2012; Chen et al., 2012; Gandomi and Haider, 2015; Kwon et al., 2014). In this framework, ‘Volume’ corresponds to the size of big data (typically multiple terabytes or petabytes). ‘Variety’, refers to the composition of the data set and, more in particular, to the structural heterogeneity in data (i.e. are the data structured, semi-structured, or unstructured). Practice shows that only a minority of the big data are structured. The ‘Velocity’ dimension refers to the dynamic nature of big data – the speed of collecting, storing and analyzing big data. Regarding this dimension, there is an increasing trend toward generating, collecting, storing, and analyzing data at high-frequency (in some sectors and applications even real-time or near to real-time). While the volume or size dimension is most discussed in the context of big data, Gandomi and Haider (2015) stress that the other dimensions are equally important. In fact, they emphasize that one should avoid focusing exclusively on one particular dimension as there may be interactions between the dimensions. For instance, the interpretation of the ‘Volume’ dimension (i.e. when can a dataset be considered big data) may very well depend on whether the data are structured or not.

Unstructured data usually require more storage and analysis capacities and better technologies than structured data. Therefore the threshold size for unstructured data will be smaller than for structured data.

Next to the three dimensions of the basic 3-V framework, also other dimensions are sometimes used to characterize big data. Gandomi and Haider (2015) and Gani et al. (2015) describe four of these dimensions: ‘Veracity’ (unreliability and impreciseness of some data sources), ‘Variability’ (similar or dissimilar data flow rates), ‘Complexity’ (few or numerous data sources), and ‘Value’ (relative value density).

All of the aforementioned characteristics impose critical challenges to the collection, storage, migration, and analysis of big data (Gandomi and Haider, 2015; Gani et al., 2015). Traditional techniques of data analysis, technologies and tools are poorly equipped to deal with these challenges and work with big data. Big data requires effective and efficient techniques and technology (as well as data organization and management) for that its potential value can be unlocked to guide decision making. Such innovative technologies need to be able to cope with the highly demanding characteristics of big data, and then in particular, the organization, storage and analysis of high volumes of fast-moving data, often from heterogeneous sources and different data types, into meaningful information. Although some new storage and computations technologies have been developed recently (for example, text mining and text analytic techniques, information extraction techniques, text summarization techniques, sentiment analysis (opinion mining) techniques, social media analytic techniques, B-tree-oriented indexing techniques, and audio and video analytic techniques), much more technological advances and analytical techniques will very likely emerge in the near future (for a state-of-the-art taxonomy of the techniques see Gani et al., 2015). A positive evolution in this respect is that new viewpoints in social science (for example, computational organizational science) are now following the developments in big data (for a good discussion of the paradigm shift for computational social sciences and big data, we refer the interested reader to Chang et al., 2014). The idea is that this will enable actors in both the public and private sector to use big data in an efficient and effective, and hence, economically feasible manner in more applications and also on a larger scale.

## **Big data and public sector: Opportunities**

In terms of the potential value of big data, there is a growing consensus among governmental stakeholders (i.e. multimedia experts, scholars, policy makers, non-governmental agencies, captains of industry) that big data applications and functionalities provide a broad range of opportunities for governments and governmental institutions worldwide (Brynjolfsson and McAfee, 2012; Chen and Zhang, 2014; Jin et al., 2015; Shaw, 2014). Resulting from this growing awareness, governments worldwide (predominantly in the US, Europe (most notably, the UK and France), Australia, Japan, Singapore, and South Korea) have announced plans and roadmaps to support the development of big data in both the public and

private sector (for an overview, see, among others, European Commission, 2010; Kim et al., 2014). Reviews of the literature (Chen and Zhang, 2014; Gandomi and Haider, 2015; Ginsberg et al., 2009; Jin et al., 2015; Morabito, 2015a) showed several interesting new and innovative applications of big data for the public sector that are already in place or that are likely to be implemented in the near future. Policy areas that have been described in the literature as having experienced considerable improvements in outcomes and services thanks to the use of big data are: the organization of traffic (Janssen et al., 2012; Lv et al., 2015), safeguarding of public security, policing (Meijer and Thaens, 2013; Meijer and Torenvlied, 2016), combatting crime and fraud (Chen and Zhang, 2014), health and well-being (Ginsberg et al., 2009), environment and sustainability (Faghmous and Kumar, 2014), transportation (Kim et al., 2014), energy (Diamantoulakis et al., 2015), smart cities (Hashem et al., 2016; Morabito, 2015a), and education (Williamson, 2016). An example of the effective use of big data in the public health sector was discussed by Ginsberg and colleagues in *Nature* (Ginsberg et al., 2009). In their article, they describe how the use of Google search queries helped in monitoring and tracking influenza-like illnesses of citizens in each region of the US so that earlier detection of influenza epidemics was possible. Positive outcomes were a more accurate prediction of the required facilities (for example, hospital beds) and vaccines, and prompt treatment of the patients. Another interesting application of big data analytics in the public sector is tax collection, an area where the call for more justice is increasingly loud. Chen and Zhang (2014) discuss how the use of big data in that area can help tax services in detecting and combatting fiscal fraud more successfully – for example, by creating profiles of people, triangulating information about people, and developing predictive models of ‘evasion taxpayers profiles’.

More generally, big data offers several advantages for public sector organizations. First, big data can help governments in making the shift from paper-filling to e-government services, for instance, through an increased integration and data flow across different PAs. While ICT is inherently driving organizations ‘paperless’, it is the combination of numerous data sources, unstructured data and data with dissimilar flow rates that make it more specific of big data. This evolution is coherent with a continuing diffusion of ICT as a tool for recording (administrative) information that can be used in a second stage. While in the past (and indeed present!) it was not uncommon for citizens to fill out multiple forms with largely the same personal information for different public service administrations, now, PAs can make use of big technologies to collect the data themselves by sharing the data sources of the other administrations or consulting on line data sources (such as Facebook<sup>®</sup> and LinkedIn<sup>®</sup>).

Second, big data can play a pivotal role in developing partnerships between governments and their citizens (Bertot et al., 2010). Whereas traditional technology provides limited possibilities to consult and inform the public about new policy instruments or services, big data technologies and infrastructure offer considerable opportunities for governments to foster civic participation in developing,

implementing and assessing policy programs. Big data applications are an important support in initiating and implementing direct online democracy, active citizen engagement, and open government initiatives (Bertot et al., 2010; Hong, 2013; O'Reilly, 2010). Margetts and Dunleavy (2013) speak of 'digital governance' which puts the interactions between humans and computers at the center of the (national and local) government business model. There is a growing interest by local governments, cities, and municipalities in innovative online tools to collect feedback from citizens and tailor public services to the citizen needs (Andrews, 2011). Mergel (2012) discussed how social media applications such as Facebook and Twitter have become widely accepted and used by the national and local governments worldwide as part of Open Government initiatives and Smart City Governance (Hoon Lee et al., 2013; Meijer, 2016). Mossberger et al. (2013) found that the use of social networks and other interactive tools in the 75 largest U.S. cities skyrocketed in recent years (with the percentage of cities adopting Facebook and Twitter increasing from respectively 13% and 25% in 2009 to 87% in 2011). Morabito (2015a) describes the example of Citysourced.com, a civic engagement software platform used by local governments and cities that offers several facilities for citizens to report and provide information to local authorities about all sorts of local problems (for example, illegal dumping, air or noise pollution, neighborhood violence, malfunctioning of street or traffic lights). In a recent opinion piece in *Public Administration Review* (O'Malley, 2014), the former mayor of Baltimore, O'Malley describes how geographic information systems (GIS) were used to collect citizen requests about city actions and services and argues that this has changed the way Baltimore is governed resulting in better administrative choices and better results. Asatryan and De Witte (2015) show for German municipalities that this form of direct democracy fosters local public government efficiency.

Third, and somewhat related to the previous advantage, big data can help PAs compile detailed and accurate profiles of citizens and using them to tailor public services to the needs and demands of the citizens (Bonsón et al., 2015; Heikkila and Isett, 2007). For instance, big data regarding citizen sentiment toward public services (most obviously, by screening the web search queries or using social media) can entail useful feedback and highlight opportunities to customize service delivery by helping employees better understand the needs of each citizen. Ho and Coates (2002) found that citizens are able to identify important aspects of government services (for example, the quality and consumer-friendliness of the provided services) that governments often ignore in the evaluation of the own performances. Incorporating these sentiments in the performance evaluation of government policies and services as well as in the implementation of changes in government policies and services also enhances the legitimacy of performance measurement as well as the transparency and the accountability toward the citizens (Bertot et al., 2010; European Commission, 2010; Lee and Kwak, 2012). The idea is that all these initiatives should also benefit citizen satisfaction with public services and governments (see Van de Walle, 2017). In addition, as discussed by Mossberger et al. (2013: 352),



the customization of information through Web 2.0 features such as RSS feeds or social networks like Facebook or Twitter may lower information costs and hence benefit the cost effectiveness of national and local government institutions and cities.

Fourth, big data can play an important role at the international level. Take, for instance, the growing interest in, and importance of, cooperation and information exchange between agencies and governments of different countries in their war on terror, the battle against tax evasion, and the international coordination of global migration. An unfortunate example of the importance of countries sharing data and information in the war on terror was the bomb attack at the Boston Marathon in 2013 which, according to several research reports, could have been prevented if Russian secret services would have shared more information with their American colleagues (Kim et al., 2014: 80). In the complex area of international tax evasion and fraud, large number of national PA databases could be integrated and shared among countries (by bi- or multi-lateral agreements) to improve fraud detection and tax evasion control (Morabito, 2015b). In the migration policy area, the importance of sharing and communicating migration data more effectively in an international context was recently demonstrated with the opening of a new Global Migration Data Analysis Centre (GMDAC) by the International Organization for Migration in Berlin (IOM, 2015).

Finally, big data can become a new source of information for public organizations for pursuing efficiency and effectiveness in their operations. Determining the efficiency of public organizations is usually a hard challenge (McConnell, 2015). Probably one of the most pervasive problems is the lack of information to determine the quality and quantity of government outputs in objective measures or figures.<sup>5</sup> Several studies (Bertot and Jaeger, 2008; Hofmann et al., 2013; Manyika et al., 2011; Mergel, 2012; Williamson, 2014) advocated that big data can provide public organizations with more detailed information about the quality and/or quantity of the governments outputs such that more adequate measures of outputs and outcomes for the public sector can be generated. For developing and implementing e-government services, for instance, Bertot and Jaeger (2008) advocated big data as a potentially valuable source of information that can help government in obtaining a clear understanding over what technologies and instruments are most efficient and effective. Interesting information could consist of measures of the awareness and engagement created by government communications – for example, the numbers of likes and comments that people have given to government posts on social media and the prevailing attitude (negative, neutral, positive) of those comments (Mergel, 2012, 2013).

## **Big data and performance management of public organizations**

Big data can also transform performance management procedures in the public sector. Most importantly, effective use of big data can boost efficiency by reducing the amount of inputs necessary for providing the current service level and/or

producing the actual output level (input efficiency) or by increasing the service and/or output level for the current input usage (output efficiency). A global survey, organized by Bloomberg Businessweek Research Services, among top managers of government agencies around the world in 2013, revealed that roughly four out of five leaders are convinced that transformations will take place in the public sector due to the use of big data (Mullich, 2013). A belief held by many managers is that, for some policy areas, big data could result in the use of entirely new management models. Take personnel performance as an example. Here, big data could be used to organize promotions, rewards or salary differentials. Big data can help Human Resource Management (HRM) departments in government institutions to identify and attract resources and talent. Performance dashboards with information on personnel performances can also be constructed and used by managers and HRM to monitor and guide the performances of personnel. In fact, HRM departments of tomorrow will use a variety of data (for example, data on working conditions, employee satisfaction and productivity) to assign tasks more optimally among divisions and employees, improve work conditions and introduce incentives that aim at improving both employee satisfaction and productivity (Brown et al., 2011). Using survey data from US local government managers, Oliveira and Welch (2013) found that social media tools are used for dissemination, feedback on service quality, participation, and internal work collaboration.

Turning to the 'institutional assessments' of public organizations, Andrews et al. (2010) discussed the importance (and the differences between) internal and external measures for assessing organizational performance. Several papers and opinion texts (O'Malley, 2014) criticize the old way of thinking about politics and governing as being largely focused on inputs and, in particular, on the question of how the resources should be allocated among the different tasks and problems. In his view, big data, and in particular, the fast collection and sharing of a variety of data, will cause a shift from an input-centric approach to an approach that focuses on outputs and outcomes. Morgeson (2014) shares this viewpoint and note that several national and local governments have already begun with shifting the focus from internal performance measures to citizen-centric measures through, among other things, the use of big data. Applications and functionalities of big data are also expected to increasingly change management models for organizing and providing public service. Government managers and heads of department could make use of performance dashboards with a large amount of operational and financial data to evaluate and compare the (cost) efficiency of departments across government agencies or different departments within governmental agencies that are performing broadly similar functions, in the spirit of benchmarking exercises.

## **Big data: Limitations and risks**

Big data does not only offer potential advantages to countries and industry, it also brings several real limitations, challenges and risks (Bertot et al., 2012; Boyd and Crawford, 2012; Picazo-Vela et al., 2012). Desouza and Jacob (2014) somewhat

roughly classified these limitations, challenges, and risks in two broad categories: (1) privacy-related problems and (2) technical difficulties. Regarding the privacy issue of big data, one particularly important question is whether the increase in use of big data may cause privacy intrusions (see Boyd and Crawford, 2012). Indeed, the activity of recording detailed individual-level information may be perceived as dangerous for citizens' intimacy and privacy. National and international legislations have the specific aim of protecting this individual right, thus acting *de facto* as a regulatory obstacle to the development of repositories for detailed information on individuals. Overall, the balance between individual rights and public interest, when concerning the sphere of personal privacy, is still an argument subjected to fierce debate (Tene and Polonetsky, 2012). As indicated by Kim et al. (2014: 81) and Yiu (2012), the line between collecting and using big data in a proper manner and sufficiently ensuring people's privacy is fine and more research should be done in order to find a good answer to this intricate question. Another issue, that somewhat relates to the previous issue is the data ownership (who owns the big data?) (Washington, 2014). Interesting cases here are recurring issues concerning data ownership with multinational social media players such as Facebook, MySpace, and Twitter. A particular problem with these global social media players is that their own rules supersede governmental regulation.

On technical hurdles, while the direction toward the use of complex, unstructured, and 'big' datasets to inform decision-making is conceptually clear, the development of systems in public organizations to handle big data effectively and efficiently is an issue that is typically far to be solved yet. Most of the challenges mainly center upon dealing with the digitization of big data, diversity of the data types, timely responding to requirements, and handling uncertainties in the data. Challenges may range from the design of storage systems that enable storing vast amounts of data, the design and implementation of collecting and processing systems which enable collecting and combining data from different sources, to the development and use of analysis techniques that enable dealing with the inherent complexity of big data. A critical point is the creation of interoperable datasets, by structurally merging systems developed by different actors, for different purposes. While this technical problem exists for both private and public organizations, it is exacerbated in the public sector, where the software used and the ability of developing ICT innovative solutions are sometimes not effective and transforming government services using ICT innovation is often complex and costly (Manyika et al., 2011; Morabito, 2015a). The existence of these technical problems should raise questions about the development of core competences within public organizations for managing big data, and make them ready for analytics. In other words, the organizations are called to assure the technical ability of working with big data, and they should not focus their attention solely on policy use of data. A white paper by Software and Information Industry Association (SIIA, 2013: 19–26) offers explicit recommendations and guidelines for policy makers, decision makers and governments to capture the potential of big data and data-driven innovation to the maximum. One such recommendation is that policymakers

should avoid establishing policies that restrain data collection and analysis. Another guideline is that policy makers should opt for flexible, open-ended rules to capture, comingle, store, and analyze big data.

There are also some threats and risks to the use of (at least some types of) big data in policy making that are more due to the inherent nature of big data. For instance, one particular threat to the use of big data that has been provided and collected using social media is that some parts of the public may not or only very limitedly participate in the information society due to the lack of knowledge, time, or facilities. Among others, Heikkila and Isett (2007) warned that even though citizens may actively participate and voluntarily provide information and feedback about delivered services, it is important for governments to keep in mind that this may only provide a partial or an incomplete picture of the experiences, criticism, and needs of the broader communities. This issue of different personality types reacting differently to the presence of social media and to social influences was nicely illustrated by Margetts et al. (2015) at the Oxford Internet Institute (OII) in several experiments (personality features that were examined include extravert, pro-self, pro-social, conscientious). An important outcome in these experiments was that whereas some types of people are typically eager to participate in social media, other types of people are less willing to participate in social media. Obviously, this impacts the quality as well as the representativeness of the big data collected by social media. Also Junqué de Fortuny et al. (2013) discuss some of the issues involved with the use of big data (missing data, miscoded data, measurement error, duplicated data, inconsistency) arguing that users of big data should be aware of the presence of such issues as well as their potential consequences – most obviously the lower quality of the big data set. An illustration of a limitation to big data use in policy decision making was discussed given by Lazer et al. (2014). In particular, as to the success story of using Google search queries for monitoring and tracking influenza-like illnesses of US citizens as presented by Ginsberg et al. (2009), they remarked that even though the use of Google search queries facilitated the monitoring it led to a persistent overestimation of flu prevalence.

A final point that is important to discuss deals with the ‘politics’ of big data – i.e. prospecting why policy-makers should use big data in their decision-making processes. This is important for several reasons. The first reason is about the possibility of innovating the way the services are developed and delivered to citizens. If, indeed, the big data allows a clearer and more precise picture of the individuals (as claimed by Pirog, 2014), then the policy-makers can better understand behaviors and preferences of citizens to tailor the specific services to them. For instance, if National Health Service can obtain timely information about people’s activities and health status, it can define an individualized set of services ready-to-use when they arrive at the hospital(s) – also by tracing an electronic set of information. Same reasoning applies to information about other spheres of public services, such as education (where Learning Analytics is indeed diffusing, see Siemens and Long 2011), elderly care, etc. In this vein, big data opens the door to a new

citizen-centricity of services' design, orientated toward a clever use of quantitative information without relying only on citizens' active involvement.

The second issue is that of more precise and robust evaluation of interventions. The promise of big data in this specific area could be seen from the perspective that '[L]arge-scale, internet derived data sets can be combined with existing traditional data from administrative procedures' (Mergel et al., 2016: 4), so that the empirical approaches used by social scientists can benefit from having a more complete set of indicators about policy outputs. Relatedly, another aspect of this new possibility stems from exploring heterogeneity of policy effects through these more integrated datasets, to concentrate the attention to the 'tails of distribution', i.e. where new collected data can help in characterizing subpopulations of citizens affected by single policies and interventions.

Third, a more extensive use of data analysis will necessarily be fostered by a continuing process of de-materialization of service delivery. Indeed, to the extent that governments will be more and more e-governments, PAs and organizations can collect users' behaviors. While the trend will conduct to some straightforward benefits (such as the reduction of bureaucracy in the intermediation of the relationships between citizens and PAs), at the same time there are indirect effects in the amount of digital information that is created in the active interaction of citizens with administrations' portals and digital infrastructures. The policy-makers will then be increasingly aware of the potential informative power of the data generated through digitally-delivered services, and they will increase the adoption of tools and instruments that are useful to trace citizens' activities and requests (as, for instance, systems of unified identification such as the digital identity cards).

## **A tentative research agenda**

To conclude, we identify some research gaps and interesting areas for future research. One promising research area with the potential to have a strong impact on big data use and research is *the study of inputs, outputs and outcomes of big data systems*. Several studies notified the need to develop efficient and effective tools to collect, store, analyze, and visualize big data (Chen and Zhang, 2014; Gandomi and Haider, 2015). The distinction between efficiency and effectiveness is important. Efficiency evaluations of big data systems focus on the input–output link, thereby asking the question how many outputs can be produced for a certain amount of inputs (for example, how many valuable information can be retrieved from big data). Evaluations of the effectiveness, on the other hand, focus on the link between the outputs and the outcomes (how accurate is retrieved information). Only a few studies have discussed the inputs, outputs and outcomes of big data systems. Among those few are Gandomi and Haider (2015) and Gani et al. (2015). Both of these studies discuss possible metrics for inputs, outputs, and outcomes of big data systems and techniques (e.g. metrics for the volume dimension of big data). Yet, more research is needed.

Another area for further research is related to the *development of a theory of how government organizations (should) adopt big data* for decision-making and organizing their actions effectively. Such theories may provide insights for managers of public organizations that can be useful for helping them in successfully implementing innovations. Some studies have made interesting attempts at studying and modeling the adoption of new innovations (such as big data) in government sector organizations. Mergel and Bretschneider (2013), for instance described a three-stage process for adopting and integrating social media in government and building communication networks for interacting with citizens and stakeholders. Broadly speaking, these three stages involve an experiment phase (informally working with social media), a regulation phase (drafting norms and regulations), and a formalization phase (the formalization of the types of interactions and new modes of communication in social media strategies and policies). Other models and critical success factors for IT-innovation adoption in government sector organizations (for instance, the Open Government Maturity Model) have been proposed and discussed by, among others, Kamal (2006), Lee and Kwak (2012). Other studies have explored new models of government practices in the era of big data and digitalization (Williamson, 2014). Nevertheless, as noted by several of these authors (e.g. Kamal, 2006), more research across different government departments and their operational settings is needed to test and further refine the model. Therefore, there is still a room for both theoretical and empirical contributions in the field. The research questions should deal with two themes: (i) to what extent big data can provide better and wider sets of information to be used by policy makers and administrators? and (ii) is a more extensive use of big data able to generate more propensity toward innovation in public services – and if yes, is this in turn leading to better results?

Another theme that warrants further research is *how new innovations such as big data affect government stakeholders* (citizens, suppliers and contractors, and politicians). A useful starting point for such research is the study of Pollitt (2011). He develops a framework for the analysis of technological change. The framework includes the effects on citizens, users of data, service providers, and other stakeholders, as well as on the wider cultural norms and beliefs. The influence of innovations on government stakeholders reveals two main trends. On one side, monitoring citizens' perspectives can favor a higher level of involvement in public decisions. This trend is not only positive because it allows engaging the citizens per se, but also because this can contrast the growing loss of trust into governments (OECD, 2013). On the other hand, big data analytics can accompany more transparency with more understanding of the underlying phenomena measured by quantitative indicators. In this sense, to the extent that the data is open and publicly available (in the open data spirit, coherent with the big data discourse), several actors can take advantage of monitoring the public organizations' activities and results. For instance, when the (big) dataset of procurement activities is made public, all the companies that supply services to the PAs can be aware of price competition; and the citizens can check the efficiency of the related expenditures. In both these cases, future research should be devoted to shed more light on

these processes of change management of public services, as well as on the effects of public value generated through these changes.

Somewhat related to the previous theme, another interesting research question to explore is *what shifts in power big data are bringing?* Aren't there any risk involved for governments (especially the local governments and the governments from less developed countries)? Isn't there any risk that big data giants such as Facebook, Google, Twitter, and others may start controlling our lives? A study report of the McKinsey Global Institute (Manyika et al., 2011) on '*Big data: The next frontier for innovation, competition, and productivity*' describes how the ownership and use of big data will become a key element of competition between enterprises and countries. Margetts et al. (2015) speak of an unruly new force in the political (but also the economic) world. Manyika et al. (2011: 6) expect that the use of big data will become a key way for leading governments and companies to outperform their peers. In particular, the belief is that leading users of big data (both in the private and public sector) who succeed to effectively capture the potential of big data will see their value and power increasing at the expense of their competitors who are more lagging in terms of using big data.

## Conclusion

There has been an increase in interest in big data technologies and related fundamental and statistical research (Chang et al., 2014; Chen et al., 2012). Illustrative is that many universities have established research centers on big data (for example, University of California at Berkeley, Columbia University, and, Eindhoven University of Technology – Jin et al., 2015). The attention of scholars is warranted given the need to establish a theory of big data. A fundamental analysis of the theory of big data would help to understand the characteristics of big data as well as to develop technologies and management models to work with big data.

In addition, a better understanding would result in clear advantages for public policy and administration. We see at least seven venues. First, by combining structured and unstructured information and data, public policy and administration will benefit from big data thanks to better services for citizens. Second, the de-materialization of procedures and bureaucracy will result in lower costs for both administrations, less personnel and lower tax rates. Also citizens will benefit thanks to fewer administrative exchanges. Third, we see big data as a solution for security issues as from the unstructured data (e.g. phone calls), data can be traced and patterns can be predicted. Fourth, in a similar vein, it might result in a solution for environmental issues as it becomes quicker and easier of keeping track of environmental problems, and providing data-driven solutions. Fifth, the current migration crisis certainly benefits from big data as administrations can easier follow (also in an unstructured way) people. Sixth, it allows policy makers to increase the citizen-centricity of services as there are more data for customizing and targeting interventions. Finally, it is possible to more precisely evaluate the interventions by exploring heterogeneity of effects via more integrated datasets.

We hope that this symposium can further contribute to the debate and fuel the knowledge of the theme. The next two papers (Agostino and Arnaboldi, 2017; Johnes and Ruggiero, 2017) provide some innovative ways of tackling the challenges ahead.

### **Declaration of conflicting interests**

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

### **Funding**

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

### **Notes**

1. We would like to thank Edoardo Ongaro, Keith Baker, Claire Dunlop, and participants of the LEER workshop 'Efficiency in Education and the use of Big Data' Leuven, 19–20 November 2015 for valuable suggestions and comments. Kristof De Witte and Tommaso Agasisti acknowledge financial support from the European Commission through the Horizon 2020 project 'Economics of Education Network' (EdEN).
2. One popular technique is the non-parametric Data Envelopment Analysis (DEA) method. In essence, DEA is a (linear programming) tool for evaluating the performance of a set of peer entities that use (possibly multiple) inputs to produce (possibly multiple) outputs. The original question in the DEA literature is how one could measure each entity's efficiency, given observations on input and output quantities in a sample of similar entities and, often, no reliable information on prices, in a setting where one has no knowledge about the 'functional form' of a production or cost function (Cherchye et al., 2007: 116–117).
3. For a recent overview, see De Witte and Lopez-Torres (2017).
4. In the remainder of the paper, we will use the terms 'big data' and 'learning analytics' interchangeably.
5. A particular problem in the measurement of government output is that price information is typically lacking or cannot be observed (Wilson, 1989). In the absence of price information, the aforementioned DEA-technique can be used to determine the efficiency of public organizations (see Cherchye et al., 2015 for a recent application of DEA in the computation of performance (productivity) ratios for public institutions).

### **References**

- Agostino D and Arnaboldi M (2017) Social media data used in the measurement of public service effectiveness: Empirical evidence from Twitter in higher education institutions. *Public Policy and Administration*. Epub ahead of print 18 December 2016. DOI:10.1177/0952076716682369.
- Andrews RW (2011) Social capital and public service performance: A review of the evidence. *Public Policy and Administration* 27(1): 49–67.
- Andrews R, Boyne GA, Moon MJ, et al. (2010) Assessing organizational performance: Exploring differences between internal and external measures. *International Public Management Journal* 13(2): 105–129.



- Asatryan Z and De Witte K (2015) Direct democracy and local government efficiency. *European Journal of Political Economy* 39: 58–66.
- Bertot JC and Jaeger PT (2008) The E-Government paradox: Better customer service doesn't necessarily cost less. *Government Information Quarterly* 25(2): 149–154.
- 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.
- Bertot JC, Jaeger PT and Hansen D (2012) The impact of polices on government social media usage: Issues, challenges, and recommendations. *Government Information Quarterly* 29(1): 30–40.
- Bonsón E, Royo S and Ratkai M (2015) Citizens' engagement on local governments' Facebook sites. An empirical analysis: The impact of different media and content types in Western Europe. *Government Information Quarterly* 32(1): 52–62.
- Boyd D and Crawford K (2012) Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication and Society* 15(5): 662–679.
- Brown B, Chui M and Manyika J (2011) Are you ready for the era of 'big data'. *McKinsey Quarterly* 4: 24–35.
- Brynjolfsson E and McAfee A (2012) Big data: The management revolution. *Harvard Business Review* 90(10): 60–68.
- Chang RM, Kauffman RJ and Kwon Y (2014) Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems* 63: 67–80.
- Chen H, Chiang RH and Storey VC (2012) Business intelligence and analytics: From big data to big impact. *MIS Quarterly* 36(4): 1165–1188.
- Chen CP and Zhang CY (2014) Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences* 275: 314–347.
- Cherchye L, De Witte K, Ooghe E, et al. (2010) Equity and efficiency in private and public education: A nonparametric comparison. *European Journal of Operational Research* 202(2): 563–573.
- Cherchye L, De Witte K and Perelman S (2015) A unified productivity-performance approach, with an application to secondary schools in the Netherlands. CES Working Paper Series 2015.19.
- Cherchye L, Moesen W, Rogge N, et al. (2007) An introduction to 'benefit of the doubt' composite indicators. *Social Indicators Research* 82(1): 111–145.
- De Borger B and Kerstens K (1996) Cost efficiency of Belgian local governments: A comparative analysis of FDH, DEA, and econometric approaches. *Regional Science and Urban Economics* 26(2): 145–170.
- Desouza KC and Jacob B (2014) Big data in the public sector: Lessons for practitioners and scholars. *Administration and Society*. Epub head of print 6 November 2014. DOI: 10.1177/0095399714555751.
- De Witte K and Geys B (2011) Evaluating efficient public good provision: Theory and evidence from a generalised conditional efficiency model for public libraries. *Journal of Urban Economics* 69(3): 319–327.
- De Witte K and Lopez-Torres L (2017) Efficiency in education. A review of literature and a way forward. *Journal of Operational Research Society* 68(4): 339–363.
- De Witte K and Moesen W (2010) Sizing the government. *Public Choice* 145(1): 39–55.
- Diamantoulakis PD, Kapinas VM and Karagiannidis GK (2015) Big data analytics for dynamic energy management in smart grids. *Big Data Research* 2(3): 94–101.

- European Commission (2010) Digitizing public services in Europe: Putting ambition into action (9th Benchmark Measurement). Prepared by Capgemini, IDC, Rand Europe, Sogeti and DTi for the DG for Information and Social Media. Available at: [http://ec.europa.eu/information\\_society/eeurope/i2010/docs/benchmarking/eGovernment\\_Benchmarking\\_Method\\_paper\\_2010.pdf](http://ec.europa.eu/information_society/eeurope/i2010/docs/benchmarking/eGovernment_Benchmarking_Method_paper_2010.pdf).
- Faghmous JH and Kumar V (2014) A big data guide to understanding climate change: The case for theory-guided data science. *Big Data* 2(3): 155–163.
- Gandomi A and Haider M (2015) Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management* 35(2): 137–144.
- Gani A, Siddiqa A, Shamshirband S, et al. (2015) A survey on indexing techniques for big data: Taxonomy and performance evaluation. *Knowledge and Information Systems* 46(2): 241–284.
- Ginsberg J, Mohebbi MH, Patel RS, et al. (2009) Detecting influenza epidemics using search engine query data. *Nature* 457(7232): 1012–1014.
- Hashem I, Chang V, Anuar N, et al. (2016) The role of big data in smart city. *International Journal of Information Management* 36(5): 748–758.
- Heikkila T and Isett KR (2007) Citizen involvement and performance management in special-purpose governments. *Public Administration Review* 67(2): 238–248.
- Ho ATK and Coates P (2002) Citizen participation: Legitimizing performance measurement as a decision tool. *Government Finance Review* 18(2): 8–11.
- Hofmann S, Beverungen D, Räckers M, et al. (2013) What makes local governments' online communications successful? Insights from a multi-method analysis of Facebook. *Government Information Quarterly* 30(4): 387–396.
- Hollingsworth B (2008) The measurement of efficiency and productivity of health care delivery. *Health Economics* 17(10): 1107–1128.
- Hong H (2013) Government websites and social media's influence on government–public relationships. *Public Relations Review* 39(4): 346–356.
- Hoon Lee JH, Phaal R and Lee SH (2013) An integrated service-device-technology roadmap for smart city development. *Technological Forecasting and Social Change* 80(2): 286–306.
- IOM (2015) IOM Opens Global Migration Data Analysis Centre in Germany. International Organization for Migration. Press Release 09/08/15.
- Janssen M, Charalabidis Y and Zuiderwijk A (2012) Benefits, adoption barriers and myths of open data and open government. *Information Systems Management* 29(4): 258–268.
- Jin X, Wah BW, Cheng X, et al. (2015) Significance and challenges of big data research. *Big Data Research* 2(2): 59–64.
- Johnes G and Ruggiero J (2017) Revenue efficiency in higher education institutions under imperfect competition. *Public Policy and Administration*. Epub ahead of print 1 August 2016. DOI: 10.1177/0952076716652935.
- Junqué de Fortuny E, Martens D and Provost F (2013) Predictive modeling with big data: Is bigger really better? *Big Data* 1: 215–226.
- Kamal MM (2006) IT innovation adoption in the government sector: Identifying the critical success factors. *Journal of Enterprise Information Management* 19(2): 192–222.
- Kim GH, Trimi S and Chung JH (2014) Big-data applications in the government sector. *Communications of the ACM* 57(3): 78–85.
- Kwon O, Lee N and Shin B (2014) Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management* 34(3): 387–394.

- Lazer D, Kennedy R, King G, et al. (2014) The parable of Google flu: Traps in big data analysis. *Science* 343(6176): 1203–1205.
- Lee G and Kwak YH (2012) An open government maturity model for social media-based public engagement. *Government Information Quarterly* 29(4): 492–503.
- Lv Y, Duan Y, Kang W, et al. (2015) Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation Systems* 16(2): 865–873.
- McConnell A (2015) What is policy failure? A primer to help navigate the maze. *Public Policy and Administration* 30(3–4): 221–242.
- Manyika J, Chui M, Brown B, et al. (2011) *Big data: The next frontier for innovation, competition, and productivity*. New York, NY: McKinsey Global Institute Report.
- Margetts H and 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(1987): 20120382.
- Margetts HZ, John P, Hale S, et al. (2015) *Political Turbulence How Social Media Shape Collective Action*. Princeton, NJ: Princeton University Press.
- Meijer AJ (2016) Smart city governance: A local emergent perspective. In: Gil-Garcia JR, et al. (eds) *Smarter as the New Urban Agenda*. Springer International Publishing Switzerland, pp. 73–85.
- Meijer AJ and Thaens M (2013) Social media strategies: Understanding the differences between North American police departments. *Government Information Quarterly* 30: 343–350.
- Meijer AJ and Torenvlied R (2016) Social media and the new organization of government communications: An empirical analysis of Twitter usage by the Dutch police. *The American Review of Public Administration* 46(2): 143–161.
- Mergel I (2012) The social media innovation challenge in the public sector. *Information Polity* 17: 281–292.
- Mergel I (2013) Social media adoption and resulting tactics in the U.S. federal government. *Government Information Quarterly* 30(2): 123–130.
- Mergel I and Bretschneider SI (2013) A three-stage adoption process for social media use in government. *Public Administration Review* 73: 390–400.
- Mergel I, Rethemeyer RK and Isett K (2016) Big data in public affairs. *Public Administration Review* 76(6): 928–937.
- Morabito V (2015a) Big data and analytics for government innovation. In: Morabito A (ed.) *Big Data and Analytics*. Springer International Publishing Switzerland, pp. 23–45.
- Morabito V (2015b) *Big Data and Analytics. Strategic and Organisational Impacts*. Morabito A (ed.). Springer International Publishing Switzerland.
- Morgeson F (2014) *Citizen Satisfaction: Improving Government Performance, Efficiency, and Citizen Trust*. Palgrave Macmillan, US, New York.
- Mossberger K, Wu Y and Crawford J (2013) Connecting citizens and local governments? Social media and interactivity in major US cities. *Government Information Quarterly* 30(4): 351–358.
- Mullich J (2013) Closing the big data gap in public sector. *Bloomberg Businessweek Research Services, Survey Report, Real-Time Enterprise*, Sep. 2013.
- Neshkova MI and Guo HD (2012) Public participation and organizational performance: Evidence from state agencies. *Journal of Public Administration Research and Theory* 22(2): 267–288.

- Oliveira GHM and Welch EW (2013) Social media use in local government: Linkage of technology, task, and organizational context. *Government Information Quarterly* 30(4): 397–405.
- O'Malley M (2014) Doing what works: Governing in the age of big data. *Public Administration Review* 74(5): 555–556.
- O'Reilly T (2010) Government as platform. In: Lathrop D and Ruma L (eds) *Open Government: Collaboration, Transparency and Participation in Practice*. Sebastopol, CA: O'Reilly Media Inc., pp.11–39.
- Organization for Economic Cooperation and Development (OECD) (2013) *Government at a Glance 2013*. Paris: OECD.
- Picazo-Vela S, Gutiérrez-Martínez I and Luna-Reyes LF (2012) Understanding risks, benefits, and strategic alternatives of social media applications in the public sector. *Government Information Quarterly* 29(4): 504–511.
- Pina V and Torres L (2001) Analysis of the efficiency of local government services delivery. An application to urban public transport. *Transportation Research Part A: Policy and Practice* 35(10): 929–944.
- Pirog MA (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.
- Pollitt C (2011) Mainstreaming technological change in the study of public management. *Public Policy and Administration* 26(4): 377–397.
- Revelli F and Tovmo P (2007) Revealed yardstick competition: Local government efficiency patterns in Norway. *Journal of Urban Economics* 62(1): 121–134.
- Sampaio BR, Neto OL and Sampaio Y (2008) Efficiency analysis of public transport systems: Lessons for institutional planning. *Transportation Research Part A: Policy and Practice* 42(3): 445–454.
- Shaw J (2014) Why big data is a big deal. *Harvard Magazine*, March–April, 30–35, 74–75.
- Siemens G and Long P (2011) Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review* 46(5): 30.
- SIIA (2013) Data-driven innovation. A guide for policymakers: Understanding and enabling the economic and social value of data. Software and Information Industry Association. Available at: [www.siia.net/Portals/0/pdf/Policy/Data%20Driven%20Innovation/data-driven-innovation.pdf](http://www.siia.net/Portals/0/pdf/Policy/Data%20Driven%20Innovation/data-driven-innovation.pdf)
- Tene O and Polonetsky J (2012) Privacy in the age of big data – A time for big decisions. *Stanford Law Review* (online) 64, 63.
- Van de Walle S (2017) Explaining citizen satisfaction and dissatisfaction with public services. In: Ongaro E and Van Thiel S (eds) *The Palgrave Handbook of Public Administration and Management in Europe*. London: Palgrave.
- Washington AL (2014) Government information policy in the era of big data. *Review of Policy Research* 31(4): 319–325.
- Williamson B (2014) Knowing public services: Cross-sector intermediaries and algorithmic governance in public sector reform. *Public Policy and Administration* 29(4): 292–312.
- Williamson B (2016) Digital education governance: Data visualization, predictive analytics, and 'real-time' policy instruments. *Journal of Education Policy* 31(2): 123–141.
- Wilson JQ (1989) *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.
- Yiu C (2012) *The Big Data Opportunity*. London: The Policy Exchange.