Politics and Governance
Open Access Journal | ISSN: 2183-2463

Volume 6, Issue 4 (2018)

Big Data Applications in Governance and Policy

Editors

Sarah Giest and Reuben Ng

COGITATIO

Politics and Governance, 2018, Volume 6, Issue 4
Big Data Applications in Governance and Policy

Published by Cogitatio Press
Rua Fialho de Almeida 14, 2º Esq.,
1070-129 Lisbon
Portugal

Academic Editors
Sarah Giest, Leiden University, The Netherlands
Reuben Ng, National University of Singapore, Singapore

Available online at: www.cogitatiopress.com/politicsandgovernance

Table of Contents

COGITATIO

Politics and Governance (ISSN: 2183–2463)
2018, Volume 6, Issue 4, Pages 1–4
DOI: 10.17645/pag.v6i4.1810

Editorial

Big Data Applications in Governance and Policy

Sarah Giest [1,*] and Reuben Ng [2,3]

[1] Institute of Public Administration, Leiden University, 2511 DP The Hague, The Netherlands;
E-Mail: s.n.giest@fgga.leidenuniv.nl
[2] Lee Kuan Yew School of Public Policy, National University of Singapore, 259772 Singapore, Singapore;
E-Mail: spprng@nus.edu.sg
[3] Geriatric Education and Research Institute, 769027 Singapore, Singapore

Submitted: 5 November 2018 | Published: 21 November 2018

Abstract
The editorial sets the scene for this thematic issue on big data applications in governance and policy. It highlights the lack of engagement in the current literature with the application of big data at the cross-section of governance of data and its utilization in the policy process and draws out aspects related to its definition and future research agenda. The contributions highlight several aspects related to big data in different contexts, such as local and national government as well as a variety of policy areas. They converge on the idea that big data applications cannot overcome existing political and structural limitations that exist in government. This leads to a future research agenda that looks at the disconnect between data production and usage as well as identifying policy issues that are more or less suitable for data analytics.

Keywords
big data; governance; policy analytics; policymaking; politics

Issue
This editorial is part of the issue "Big Data Applications in Governance and Policy", edited by Sarah Giest (Leiden University, The Netherlands) and Reuben Ng (National University of Singapore, Singapore).

1. Introduction

Recent literature has been trying to grasp the extent as to which big data applications affect the governance and policymaking of countries and regions (Boyd & Crawford, 2012; Giest, 2017; Höchtl, Parycek, & Schöllhammer, 2015; Poel, Meyer, & Schroeder, 2018). The discussion includes the comparison to e-government and evidence-based policymaking developments that existed long before the idea of big data entered the policy realm. The theoretical extent of this discussion however lacks some of the more practical consequences that come with the active use of data-driven applications. In fact, much of the work focuses on the input-side of policymaking, looking at which data and technology enters the policy process, however very little is dedicated to the output side. In short, how has big data shaped data governance and policymaking? The contributions to this thematic issue shed light on this question by looking at a range of factors, such as campaigning in the US election (Trish, 2018) or local government data projects (Durrant, Barnett, & Rempel, 2018). The goal is to unpack the mixture of big data applications and existing policy processes in order to understand whether these new tools and applications enhance or hinder policymaking.

Existing research is split regarding the usefulness of big data in the policy realm. Some are convinced that there is nothing new in the way data is being used—even if it is *big* data. This argument is in reference to the large administrative datasets that government has handled prior to the big data idea and the technological shift that came with the introduction of computers and increasingly refined software to utilize data (Connelly, Playford, Gayle, & Dibben, 2016). Others however see a shift at the scale of the Industrial Revolution (Richards & King, 2014), due to the type and speed of information being available. Since there is a variety of big data applications and governance systems, it is difficult to find one answer to the question whether big data will permanently alter the policymaking process. With this thematic issue we aim to contribute to this discussion by highlighting applications in a variety of contexts to show that they come to a common conclusion: there is benefit to using big data in the policy realm, however (1) a more nuanced look at ongoing applications reveals a complex picture of politics entering the process, and (2) contextual factors, such as the level of government, the policy field and the hierarchical structure affect data utilization. In other words, big data applications cannot overcome existing political and structural limitations that exist in government. This finding might be a less exciting one, but is a cautionary warning to those governments that portray big data as numbers-only, neutral information that can solve a variety of issues.

The following section gives an overview over the definition of big data in the governance and policymaking literature and is followed by a summary of the contributions to this thematic issue. The editorial concludes with ideas for future research.

2. Big Data in Governance and Policy

The concept of big data is vague and has not been clearly defined (Connelly et al., 2016). The articles in this issue converge on a definition that acknowledges the different forms in which big data can appear in the policy process. For example, Durrant et al. (2018) consider administrative data as a form of big data, because it is exhaustive, highly granular, large and found and repurposed, rather than made. Trish (2018) also focuses on the use of administrative and performance data as part of a long-standing evidence-based policy movement in the US government. Longo and Dobell (2018) acknowledge census data as big data, and focus in their paper on its velocity and variety as a foundation for policy analytics. Ng (2018) defines big data as unstructured data that a city produces such as video, audio, sensor data, citizens' conversation online and social media. This zooms in on the volume and veracity of the data available.

There are two ways of understanding the use of big data in government. One is to look at the governance of big data, which includes the handling and regulation of data. The other perspective is to focus on the utilization of big data for specific policy problems. In this issue we collapse both into the idea of big data in governance and policy-making based on the assumption that they are intrinsically linked. This linkage is visible when data regulations prevent the collaboration of government units for addressing cross-cutting issues (Durrant et al., 2018). Another intersection of big data for policymaking and the governance of it are the challenges highlighted by Trish (2018) around public scrutiny of the information being used by government. Here, questions are raised around how the data is governed in terms of its transparency and values and as well as how this information is used to make decisions around public policy.

3. Contributions to This Thematic Issue

Longo and Dobell (2018) begin the thematic issue with an overview of theoretical and applied work in policy analytics. They define policy analytics as a modification to the traditional policy analysis approach and position this idea in a wide variety of literature while giving practical examples of its application. By looking at the emergence of policy analytics within the policy sciences, they find that new ways to analyze policies is much more than just data analysis. Based on a review of recent literature, they show that the promises that data-driven applications make is met with the complexity of policy decisions. This intersection is where less researched, but interesting questions are raised in terms of whether the policy environment is too complex for even advanced policy analytics to contribute or whether the effects of one policy decision is so diffused in a variety of sectors and governmental levels that the effect of policy analytics is hard to grasp. Longo and Dobell (2018) conclude with a matrix for the applicability of policy analytics across scale (from local to global) and complexity (from uncertain to certain). This illustration shows that policy analytics can best support local problems that have a degree of certainty, such as monitoring, implementation and enforcement.

In their contribution, Durrant et al. (2018) pick up on the idea of local big data applications and use participatory action research to observe activities of identifying, integrating and analyzing multiple and diverse forms of data. Based on this, they theorize about the social contexts of both data production and policymaking to better understand the boundaries and barriers to big data for policy. In their work, Durrant et al. (2018) find that the context for data applications is deeply value-laden and political, which leads them to draw the following conclusions. First, there is an absence of data sufficiently relevant for addressing specific policy questions. In other words, the questions being raised in, for example health policy, could not be answered with the data being routinely collected and made available to the responsible agency. Second, the data being collected is largely used for service administration and audit rather than tackling underlying issues, such as reasons for low service take-up. Finally, the cost of providing data is greater than the perceived benefit. This has to do with having to establish the validity of data access requests by different authorities, data holders and project teams being involved. Taken together, Durrant et al. (2018) conclude that the insights from available data are not always actionable in the local context due to the factors mentioned above and that caution should be exercised when it comes to which questions can be asked of big data.

Trish (2018) focuses on the data use at national level in the US. She analyses three cases from the Obama Administration: microtargeting in electoral campaigns, performance management in government and signature drone strikes. Whereas these applications are highly technical in nature, the paper shows that, similar to the previous two analyses, underlying assumptions and power relationships impact the usefulness of data. In fact, decisions are often made based on incomplete information and Trish (2018) cautions the uncritical use of data by having efficiency as a foundation for such decisions. She finds that there is limited public scrutiny in combination with an undercurrent of market-based influence. Looking ahead, Trish (2018) concludes that using big data in this way reinforces existing biases of society and gives decisions an appearance of objectivity, which is not warranted based on the type of data that is being used. With this assessment, she underpins the previous findings that big data has a role to play, but the information drawn from the data has to be used with caution in terms of their completeness, applicability and the type of question they are supposedly answering.

Finally, Ng (2018) provides a case study of Singapore's big data applications in governance and policy that are enabled by cloud computing adoption. He distills five key factors that drive the use of big data in public management and policy: (1) public demand for big data applications; (2) focus on whole-of-government policies and practices; (3) restructuring of technology agencies to integrate strategy and implementation; (4) creating the Smart Nation Platform; (5) purpose-driven big data applications especially in healthcare. Taking lessons from Singapore, he concludes that other countries can promote regulatory sandboxes to experiment with policies that proactively manage novel technologies and business models that may radically change society, and establish more public-private partnerships to co-innovate on challenges.

4. Concluding Remarks and Ideas for Future Research

This thematic issue raises a non-exhaustive list of issues linked to big data in governance and policy. The contributions shed light on a range of factors that have been partially overlooked in current research on the topic. In particular, all papers converge on the idea that policymaking is a complex process in which data analytics is one factor that might have positive, negative or no effect at all. In fact, the papers highlight that the positive effect is over-valued, which leads to decisions being made based on incomplete evidence (Trish, 2018) or irrelevant information regarding the problem at hand (Durrant et al., 2018). The contributions further give the sense that the production and use of data remain two separate processes, which means that the data are not answering the questions linked to specific policy issues. This disconnect leads to data-based evidence that is incomplete, not actionable or even confusing from the perspective of policymakers looking for answers. Hence, a future research question would be how this disconnect comes about and how policy issues can inspire data collection rather than existing data informing solutions for policy problems.

Another issue raised in the contributions for this thematic issue, is the complexity of policymaking, which cannot be simplified by more data. In fact, data has not shown to be as disruptive to existing processes due to long-standing political, hierarchical and procedural structures. As Longo and Dobell (2018) point out, the context in which big data tools are applied matters in terms of its complexity and scale. Looking ahead it raises the question for big data and policy research, whether data use is particularly applicable to activities, such as monitoring and unfit for more complex issues, such as community health services (Durrant et al., 2018). Essentially, data have to achieve a purpose. As Ng (2018) concludes, a data project without clear policy goals careens into disillusionment, and negatively impacts the perception of data in the policy process.

Finally, the contributions agree that data use in policymaking is not a linear process where data is analyzed and then information fed into the policy cycle. In fact, barriers to data use occur in unlikely situations, such as the sharing of data with private companies who then deny access to it for integration (Durrant et al., 2018) or the need for qualitative statements next to predictive models due to unexpected outcomes (Longo & Dobell, 2018). This points towards questions of trust in the process of sharing and using certain type of data. Looking ahead, research institutions, such as universities, could play a unique role by bringing together public and private organizations to achieve mutually beneficial outcomes. Ng (2018) describes such a formalized approach where a Memorandum of Understanding (MOU) is signed among multiple private and public entities to co-create solutions for complex societal challenges.

Acknowledgements

The editors would like to acknowledge and thank the contributing authors and the anonymous reviewers of this thematic issue for their work.

Conflict of Interests

The authors declare no conflict of interest.

References

Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, *15*(5), 662–679.

Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, *59*(C), 1–12.

Durrant, H., Barnett, J., & Rempel, E. S. (2018). Realising the benefits of integrated data for local policymaking: Rhetoric versus reality. *Politics and Governance*, *6*(4), 18–28.

Giest, S. (2017). Big data for policymaking: Fad or fast-track? *Policy Sciences*, *50*(3), 367–382.

Höchtl, J., Parycek, P., & Schöllhammer, R. (2015). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, *26*(1/2), 147–578.

Longo, J., & Dobell, R. (2018). The limits of policy analytics: Early examples and the emerging boundary of possibilities. *Politics and Governance*, *6*(4), 5–17.

Ng, R. (2018). Cloud computing in Singapore: Key drivers and recommendations for a smart nation. *Politics and Governance*, *6*(4), 39–47.

Poel, M., Meyer, E. T., & Schroeder, R. (2018). Big data for policymaking: Great expectations, but with limited progress? *Policy and Internet*, *10*(3), 347–367.

Richards, N.bM., & King, J. H. (2014). Big data ethics. *Wake Forest Law Review*, *49*, 393–432.

Trish, B. (2018). Big data under Obama and Trump: The data-fueled U.S. presidency. *Politics and Governance*, *6*(4), 29–38.

About the Authors

Sarah Giest is an Assistant Professor at the Institute of Public Administration, Leiden University. She is conducting research on urban policymaking, specialising in networking among various governmental and non-governmental stakeholders and the policy design connected to that. Sarah is currently working on several papers and projects connected to big data use by policymakers together with researchers from Computer Science with the goal of unpacking the way data information enters the policy process in the city context.

Reuben Ng spent 16 years in government, consulting, and research. In government, he was in the Prime Minister's Office of Singapore driving evidence-based policy-making through data analytics, and Smart Nation strategies. In consulting, he co-built the Advanced Analytics practice at a top firm and implemented complex analytics projects across industries and functions. In research, he is an expert in quantitative social sciences, social gerontology, and credited with creating innovative techniques to measure societal perceptions/stereotypes that are applied to policy, and program evaluation. Reuben trained as a behavioural scientist at NUS, Oxford and Yale where he was Singapore's first Fulbright Science and Technology Scholar.

Politics and Governance (ISSN: 2183–2463)
2018, Volume 6, Issue 4, Pages 5–17
DOI: 10.17645/pag.v6i4.1561

Article

The Limits of Policy Analytics: Early Examples and the Emerging Boundary of Possibilities

Justin Longo [1,]* and Rod Dobell [2]

[1] Johnson Shoyama Graduate School of Public Policy, University of Regina, Regina, S4S 0A2, Canada;
E-Mail: justin.longo@uregina.ca
[2] Centre for Global Studies, University of Victoria, Victoria, V8P 5C2, Canada; E-Mail: rdobell@uvic.ca

* Corresponding author

Submitted: 1 May 2018 | Accepted: 19 September 2018 | Published: 21 November 2018

Abstract

Policy analytics has emerged as a modification of traditional policy analysis, where the discrete stages of the policy cycle are reformulated into a continuous, real-time system of big data collection, data analytics, and ubiquitous, connected technologies that provides the basis for more precise problem definition, policy experimentation for revealing detailed insights into system dynamics, and ongoing assessment of the impact of micro-scale policy interventions to nudge behaviour towards desired policy objectives. Theoretical and applied work in policy analytics research and practice is emerging that offers a persuasive case for the future possibilities of a real-time approach to policymaking and governance. However, policy problems often operate on long time cycles where the effect of policy interventions on behaviour and decisions can be observed only over long periods, and often only indirectly. This article surveys examples in the policy analytics literature, infers from those examples some characteristics of the policy problems and settings that lend themselves to a policy analytics approach, and suggests the boundaries of feasible policy analytics. Rather than imagine policy analytics as a universal replacement for the decades-old policy analysis approach, a sense of this boundary will allow us to more effectively consider the appropriate application of real-time policy analytics.

Keywords

adaptive management; agency; big data; data analytics; governance; nested institutions; nudging; policy analysis; policy analytics

Issue

This article is part of the issue "Big Data Applications in Governance and Policy", edited by Sarah Giest (Leiden University, The Netherlands) and Reuben Ng (National University of Singapore, Singapore).

> If optimal control theory becomes fully operational in economics in the next few years...economists will have at their disposal a mathematical supertool that...actually tells you what policy to use...the best possible timing and dosage for each available policy remedy. (Business Week, 1973, p. 74)

1. Introduction

Policy analytics has emerged in recent years as a modification of the traditional policy analysis approach, where the discrete stages of the policy cycle are being reformulated into a continuous, real-time system of big data collected from ubiquitous, connected technologies, assessed using advanced data analytics. Technological developments now provide policymaking with access to massive amounts of real-time data about policy problems and system conditions. When coupled with growing capacities in data analytics, policy analytics provides a basis for more precise problem definition, detailed insights into system dynamics, and ongoing assessment of the impact of micro-scale policy interventions to nudge behaviour towards desired policy objectives (Daniell, Morton, & Insua, 2016; De Marchi, Lucertini, & Tsoukiàs, 2016; Höchtl, Parycek, & Schöllhammer, 2016; Kitchin, 2014; Lazer et al., 2009; Mergel, Rethemeyer, & Isett, 2016; Tsoukias, Montibeller, Lucertini, & Belton, 2013).

Policy analytics presents a mix of technology and expertise that could result in important advances in the science of policymaking (Giest, 2017). However, despite some early successes and enthusiasm for the possibilities of policy analytics, a number of questions and barriers to their use have emerged, principally issues related to privacy risks, data biases, and the need to clarify the relationship between the technocratic accuracy of policy analytics, and the challenges of decision-making in a diverse democracy (Noveck, 2018). Our focus here is on a specific concern that remains underexplored: to identify where the strengths of policy analytics live up to its billing, consider what the range of plausible applications is, and begin to assess the limits of policy analytics for addressing public policy problems. Our guiding research question asks what types of policy problems are amenable to 'fast' feedback control systems facilitated by big data and analytics, and which require a deeper, patient, 'slower' more deliberative approach to problem definition, analysis, decision-making, implementation, and evaluation (Kahneman, 2011). To pursue this question, we undertake a survey of the literature in policy analytics theory and practice, deriving from that the features of policy problems and their settings that characterize the range of policy issues to which policy analytics can reasonably be applied, leading towards a sketch of the boundaries of policy analytics. Rather than imagine policy analytics as a universal replacement for the decades old policy analysis approach, understanding this boundary will allow researchers and practitioners to more effectively consider the appropriate application of a real-time policy analytic approach. Our claim is that policy analytics complements and supports democratic deliberation and civic engagement; with agreement on operational objectives, policy analytics built on big data makes effective feedback control feasible.

We start by defining what we mean by policy analytics as distinct from policy analysis, sketch the emergence of the technological possibilities that have given rise to policy analytics and outline some concerns that have emerged. We next present a scan of recent policy analytic examples, leading to the identification of some characteristics of policy issues that are amenable to a policy analytics approach and—by extension—some types of policy issues that are not suitable to a continuous, real-time system of big data and data analytics, concluding with some guidance as to when policy analytics might be considered an appropriate approach. This boundary around the possibilities of policy analytics should supplement the broader need to consider the appropriate place for a policy analytic approach in the context of representative and deliberative democracy, social justice and equity considerations, social diversity, and citizen privacy rights, concerns that should temper any unexamined enthusiasm for policy analytics.

2. The Emergence of Policy Analytics within the Policy Sciences

The modern policy analysis movement is based on an integrated, multidisciplinary approach to the study of public problems and the development of rational solutions based on careful analysis of evidence (Lerner & Lasswell, 1951). Decisions based on the best available evidence and rigorous analysis should be better positioned to address public problems than those based on anecdote, unsupported belief, or inaccurate data (Quade, 1975). From those origins, policy analysts have traditionally been tasked with precisely defining policy problems, collecting and analyzing data and evidence, supporting political decision-making with advice, guiding faithful implementation of those decisions, and objectively overseeing the evaluation of how effective those policy interventions were.

During the first quarter century of the policy analysis movement, quantitative techniques became staples of the theory and practice of policy analysis (Quade, 1980; Radin, 2000). Despite these significant advances and successes, debates over the perceived and proposed role of policy analysis have persisted in the profession's later years (Dryzek, 1994; Stone, 1988). While technical, empirical, quantitative policy analysis became increasingly sophisticated during the 1970s, and since, high-profile failures and the perceived inability to solve complex public problems exposed the limits of positivist policy analysis (May, 1992). Critics of positivism argued that the attempt to model social interactions using mathematical models was misguided (Amy, 1984), that policy analysis was much more than data analysis (Meltsner, 1976; Wildavsky, 1979), and that positivism was fundamentally incapable of dealing with complex problems in a democracy (Fischer, 1995). A "malaise...of the policy sciences" crept into the discipline as its positivist, neo-classical economics orientation seemed incapable of understanding human behaviour, accommodating the democratic expectations of citizens, or remedying the increasing complexity of policy problems (Deleon, 1994, p. 82). The positivist policy analysis hegemony was also undermined by limitations in data availability and the tools of analysis (Morgan, Henrion, & Small, 1992). Analysts inclined towards quantitative methods longed for even more robust data, greater computational power, and the development of more technically sophisticated policy analysis throughout government and wider policy circles (Morçöl, 2001). Some of those goals appear to have been attained in the digital era, with the growth of big data arising from the ubiquitous deployment of networked computing devices throughout society and increased data analytic capacity to manage the resulting flood of data.

Definitions of 'big data' abound (Dutcher, 2014; Fredriksson, Mubarak, Tuohimaa, & Zhan, 2017; Ward &

Barker, 2013), with most focusing on its characteristics—especially the large *volume* of data, its continuous flow at high *velocity*, and the *variety* of data available—and others pointing to the complexity of combined data sets and their value in revealing previously undetectable patterns. What emerges, however, from the policy analytics literature is a frequent conflation of 'big data' with 'large' data collections such as a census. While this reflects the current state of the art, our concept of big data draws especially on the velocity and variety (and, consequently, the large volume) of data as the foundation for a policy analytic approach that centres on a real-time understanding and interaction with the policy environment.

With the emergence and expansion of the Internet and the range of digital technologies that have been deployed in recent years, analysts now have access to a wide range of policy-relevant big data. These technologies and their users generate a variety of signals through devices like mobile smartphones, Internet of Things (IoT) devices, personal wearables, electronic transaction cards, *in situ* sensors, web search and web traffic, and social media. Massive amounts of data are now generated continuously through the daily activities of individuals, from their interactions with web services and social media platforms, purchasing behaviour and transportation choices revealed through electronic transaction cards, movement and interaction captured through mobile smartphones and wearables, behavioural choices measured through IoT consumer products, a range of measurements captured by sensors, satellite remote sensing, counters, and smart meters, and the interactions of people and devices with control technology. The accumulation of these signals, and associated metadata such as geolocation information and time stamps, results in a previously unimaginable amount of data, precisely measured from multiple perspectives, and captured in real time. Advances in data storage technologies now make it possible to preserve increasing amounts of data, and faster data transfer rates allow for cloud computing at low cost. We can now capture, store, and process data—in volumes previously unimaginable, from ubiquitous sources, with continuous flow, observed through multiple channels—and have increased capacity to manage, analyze, and understand these new data sources (Lazer et al., 2009). Not only has the volume of data and our ability to analyze it changed. The same technologies that allow for real-time data capture from the field provide a mechanism to communicate policy signals outward to actors, agents, and those devices, serving again to gather further data that measure reaction to those signals. With the stages thus joined up, policy formulation can be connected with implementation and evaluation processes in a continuous and real-time cycle of ideation, experimentation, evaluation, and reformulation (Pirog, 2014). New digital tools, platforms, and the data they generate allow for a seamless linking of the discrete stages of the policy cycle into a continuous, real-time, feedback cycle of problem identification, tool modification, system monitoring, and evaluation. This technology revolution offers the potential to revive and extend the positivist tradition in policy analysis and offer improved support for policymaking through an approach we call 'policy analytics'.

To be certain, there are competing conceptualizations of what policy analytics implies (Daniell et al., 2016; De Marchi et al., 2016; Tsoukias et al., 2013). While referred to *inter alia* as 'big data' applied to public policy and administration (Einav & Levin, 2014a; Giest, 2017; Höchtl et al., 2016; Kim, Trimi, & Chung, 2014; Kitchin, 2014; Mergel et al., 2016), 'computational social sciences' (Lazer et al., 2009), and 'policy informatics' (Johnston, 2015), the term policy analytics is used here to emphasize the combination of new sources and forms of policy-relevant big data with the use of new analytic techniques and capacity to affect policymaking throughout the entire policy cycle. Some definitions stretch the definition of 'big data' to include traditional—albeit very large—government 'large data' collections such as censuses, taxation data, social security records, health information, and survey data (Daniell et al., 2016). Some perspectives emphasise this supplementing of large data with big data, where datasets are linked with the aim of identifying previously undiscovered patterns and correlations at the problem identification and analysis stages (Höchtl et al., 2016; Janssen & Kuk, 2016a). Others focus on high volume real-time big data, combined with highly structured administrative large data, for deriving insights for operations and public service delivery (Joseph & Johnson, 2013; Mergel et al., 2016).

The harvesting of big data, coupled with advances in technology and scientific developments for managing that data, emerged first in the private sector under the heading 'business analytics', with analytics serving as an umbrella term for statistical methods and approaches including statistics, data mining, machine learning, business intelligence, knowledge management, decision support systems, operations research, and decision analysis. Key to the development of business intelligence was that this intelligence was useful if it led to action that was immediate and the impact of that action measurable (Longo, 2018; McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012). When eventually applied to public policy problems, this led to the concept of 'policy analytics' denoting the development and application of data analytic skills, methods, and technologies, supporting stakeholders with meaningful and informative analysis at any stage of the policy cycle (De Marchi et al., 2016; Tsoukias et al., 2013). Pirog (2014) envisions the extension of previously developed quantitative methods through the linking of government administrative records, data from natural science fields such as biology and neuropsychology, and geospatial data ushering in a dramatic advance in policy research. Giest (2017) gives examples from different policy domains—health, education, climate change, and crisis management—and identifies a mix of data, technology, and expertise that could result in important advances in the science of policymaking. Thus, based on the literature that has emerged to date from both business analytics and policy analytics, we define policy analytics as the use of new sources and forms of policy-relevant big data combined with advanced analytics techniques and capacity, taking advantage of ubiquitous communication methods to reduce the time delay amongst stages of the policy cycle, aimed at better addressing public problems.

In adopting the tools of policy analytics, governments are mirroring the actions of private sector firms that use big data to better understand people's behaviour. Examples include encouraging users to return to a webpage, click on an ad, buy a product and a subsequent product, purchase a service, or watch a movie because they watched a similar one (McAfee et al., 2012). Data analytics can also be used to judge who is a worthy credit risk, who would be a good person to hire, and who would make an ideal mate (Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016; Tufekci, 2014). Despite these early successes and enthusiasm for the possibilities of policy analytics, a number of questions and barriers to their use have emerged that should temper any unexamined enthusiasm (Noveck, 2018). Among these are concerns over privacy and security of citizens' data (Kim et al., 2014), proper and efficient permissioning to facilitate use by public servants (Welch, Hinnant, & Moon, 2005), weak data skills among public servants and a reliance on external consultants and contract data analysts (Dunleavy, Margetts, Bastow, & Tinkler, 2006), faulty analysis where strong correlations are valued over preliminary causal explanations (Harford, 2014), questions about big data representativeness as new digital divides emerge that undermine the possible democratizing effects of policy analytics (Longo, Kuras, Smith, Hondula, & Johnston, 2017), establishing the provenance of big data so that stakeholders and decision makers can understand where the evidence came from (Javed, McClatchey, Khan, & Shamdasani, 2018), opacity in policymaking by algorithm (Kitchin, 2017; Mittelstadt et al., 2016; Pasquale, 2015), bias in algorithms and machine learning (Koene, 2017), an over-reliance on data for decision-making in situations where values are important (Majone, 1989; Shulock, 1999), and its inverse, ignoring data in decision-making (Harsin, 2015; Tingling & Brydon, 2010).

Policy analytics represents a persuasive combination of advanced digital technology and modern behavioural science. But it has emerged alongside volatile and untrustworthy information and communications technologies reshaping shifting perceptions and redirecting changing beliefs, driving the evolving preferences that must be reflected in contested metrics for signalling social welfare and community wellbeing. In assessing this challenge, it is necessary to consider what kinds of public reasons can legitimately support the authoritative exercise of delegated public power in a political setting marked by a lack of consensus within a divided society. As the potential dangers of the big data industry begin to be revealed and slowly understood (Persily, 2017), the question that must be asked of government is whether the benefits of policy analytics outweigh the potential downsides (Boyd & Crawford, 2012). This challenge is, of course, just one facet of the broader social question of what it means to retain meaningful human control of technocratic instruments, including autonomous and intelligent systems, in a world where the exercise of human agency is increasingly distanced from consequences and individual responsibility.

3. Policy Analytics in Practice

Policy analytics can take a range of approaches. Perhaps the simplest, first line of analysis lies in social media monitoring or 'social listening' to analyze and respond to citizen's preferences, experiences, articulated values, and behaviours (Charalabidis, Loukis, Androutsopoulou, Karkaletsis, & Triantafillou, 2014; Grubmüller, Götsch, & Krieger, 2013; Prpić, Taeihagh, & Melton, 2015). Social listening involves searching and monitoring social media for words, phrases, hashtags, or mentions of government accounts or persons. This approach is becoming increasingly popular with governments seeking to gauge public perception and better appreciate why citizens have the attitudes they do and how these attitudes change over time (Longo, 2017; Paris & Wan, 2011). Further analysis can centre on the assessment of sentiment and meaning, clustering opinion to reveal network properties and make sense of public opinion (Till, Longo, Dobell, & Driessen, 2014).

Venturing deeper, predictive analytics can serve as an input into framing a policy problem before it is apprehended as such, indicating where a need is being unmet, or where an emerging problem might be countered early. As a big data analytics form of forecasting (Sims, 1986) now referred to as nowcasting (Choi & Varian, 2012), predictive analytics is based on the argument that analysis of past performance can reveal a probable outcome that can be expected from continuing to pursue the same approach (i.e., doing nothing). Some recent initiatives show the possibilities for success, for example in reducing administrative failures (Behavioural Insights Team, 2012) and understanding social dynamics (Bond et al., 2012). The combination of digital signals and new analytic techniques can help in understanding and predicting behavior in contexts such as crime (Chan & Bennett Moses, 2015), energy use (Zhou & Yang, 2016), migration (e.g., the use of email, social media, web search, and geolocation have been used to infer migration flows; see Gerland, 2015; Raymer, Wiśniowski, Forster, Smith, & Bijak, 2013; Verhulst & Young, 2018), urban planning (Kitchin, 2014), and public health (Khoury & Ioannidis, 2014; Murdoch & Detsky, 2013).

Policy experimentation builds on the idea of policy incrementalism (Lindblom, 1959), with a long history of examples of trials, experiments, and pilots of varying scale and precision, and a renewed enthusiasm in jurisdictions from the United Kingdom (Breckon, 2015) to Canada (Monafu, Chan, & Turnbull, 2018). Real-time experimental policy analytics takes advantage of new big data sources, coupled with data analytics techniques, bringing together all the discrete stages of the policy cycle into one continuous process. While a policy problem is being observed, interventions would also be underway using the same devices used to collect the data, with their impact on the problem becoming part of the evidence base for further modifying the policy variables. These further modifications would also be observed for their impact, as the system response to the policy intervention moved closer to the policy target or equilibrium (Esperanza & Dirk, 2014). An intriguing application of policy analytics from transportation management can be seen in the evolution from high-occupancy vehicle (HOV) lanes to high-occupancy smart toll (HOST) lanes (Longo & McNutt, 2018).

Shi, Ai and Cao (2017) argue that some policy analytic methods are better suited to particular stages of the policy cycle than others, and provide several examples to support their claim. Cognitive mapping, text mining, and understanding public attitudes through geo-specific Google search-query data (Lee, Kim, & Lee, 2016) are applicable to the agenda-setting phase, participatory planning in the decision phase, and remote sensing, smart metering, or participatory GIS to monitoring and evaluation phases. Decision support systems to collect, manage, and analyze data (e.g., a space-air-ground big data traffic system that includes people, vehicles, and road conditions using data from satellite sensing, aerial photography, aerial drone sensors, cameras, transponders, and smartphones) can support overall transportation policy implementation, law enforcement, and emergency response (Xiong et al., 2016). A groundwater web portal that combines legacy data, community-sourced groundwater information, and government open data provides real-time information to the public, and tools for data querying and visualization to support decision-making and community engagement (Dahlhaus et al., 2016). A big data archive covering more than 43 million soldiers, veterans, and their family members provides a foundation for the examination of the causes and consequences of PTSD (Vie, Griffith, Scheier, Lester, & Seligman, 2013).

In some cases, policy evaluation can be undertaken using policy analytics. Lu, Chen, Ho and Wang (2016) analyze 2 million construction waste disposal records to assess the disparity between public and private operator performance, with contractors operating in public projects performing better than those in private projects. In transportation management, cases from the Netherlands and Sweden show that automated smartcard and vehicle positioning data provide for better understanding of passenger needs and behaviours, system performance, and real-time conditions in order to support planning and operational processes (Van Oort & Cats, 2015).

Participatory policy analytics can take the form of sentiment analysis, mined from Internet content including social media, used to gauge how the public values alternative outcomes. Beyond simplistic exercises such as counting 'likes' and 'mentions', the example of mining Yelp restaurant reviews as a supplement (and potential replacement) for public health inspections (Kang, Kuznetsova, Luca, & Choi, 2013) shows that mining of large volumes of text contributions from citizens concerning government policies can extract opinions and knowledge useful for policy purposes (Maragoudakis, Loukis, & Charalabidis, 2011).

Poel, Meyer and Schroeder (2018) present the results of a recent project that scanned for big data policymaking examples, noting the heightened interest in big data for policymaking in recent years, though acknowledging that there are still few good examples available. They analyze 58 data-driven cases, with a focus on national and international policy initiatives, and highlight persistent challenges: data representativeness, validity of results, gaps in citizen engagement, and weak data analysis skills. While most examples do not tread on personally identifiable data, privacy protection remains a concern due to re-identification/de-anonymization risks (de Montjoye, Hidalgo, Verleysen, & Blondel, 2013; Narayanan & Shmatikov, 2008). More generally, using big data for policy purposes revives concerns about technocracy, technoscience, policy-based evidence making, and the influence of lobby groups. The most prominent area Poel et al. (2018) identify centres on government transparency initiatives supported by the publication of open data on procurement, having the objective of revealing government corruption. A smaller number of initiatives focus on operational policy areas such as budgeting, economic and financial affairs, and transportation. Remaining initiatives cover policy areas such as health, education, research, justice, and social affairs. Almost half of the initiatives scanned focus primarily on the early stages of the policy cycle (e.g., sentiment analysis via Twitter to support agenda setting and problem analysis), with others supporting policy design, implementation, and monitoring. Observing traffic patterns via sensors and mobile phone data, and using administrative data to monitor transportation and environmental policies, were also highlighted. However, as most of the projects scanned in Poel et al. (2018) use data formats such as spreadsheets, and data analysis is limited to descriptive statistics or occasional visualizations with few examples of techniques such as machine learning or algorithmic response, the boundary in this survey between 'large data' and 'big data' appears fluid.

Schintler and Kulkarni (2014) review the range of arguments for and against the use of big data in policy analysis, and offer examples to illustrate some of the positive features. The 'Billion Prices Project' uses web-sourced price information from retailers across multiple countries and sectors to generate daily estimates of inflation, providing a real-time price index as opposed to the periodic figures produced by national statistical agencies (Cavallo & Rigobon, 2016). The 'Global Forest Watch' project processes hundreds of millions of satellite images as well as data from people on the ground to generate real-time estimates of tree loss that are more precise than those produced from other approaches (Hartmann et al., 2018). The near real-time data are available freely online, and have been used to measure global deforestation rates, detect illegal clearing activity and burning, and monitor corporate sustainable forestry commitments.

Daniell et al. (2016) point towards examples of policy analytics for formulation or delivery in the areas of health resource allocation (Aringhieri, Carello, & Morale, 2016), sentiment analysis and opinion mining (Alfaro, Cano-Montero, Gómez, Moguerza, & Ortega, 2016), using behavioral information to encourage energy efficiency, precision government services (Hondula, Kuras, Longo, & Johnston, 2017), identifying social service and public information 'deserts' (Entwistle, 2003), and promoting smart cities (Kumar, Nguyen, & Teo, 2016). Additional examples are being tested, and stand as potential opportunities for applied policy analytics, from using smart electricity meters to incentivize conservation behaviour and reduce peak-load demand (Blumsack & Fernandez, 2012; Newsham & Bowker, 2010), to possibilities such as creating on-demand local public transportation services (Murphy, 2016). The Joint Statistical Research Program of the US Internal Revenue Service enables studies that use long panels of tax returns to observe individuals over time with a view to revealing potential policy initiatives (Jarmin & O'Hara, 2016).

The principles of nudge theory are being applied in dynamic ways that take advantage of the powerful devices ubiquitously moving around us to measure the environment, along with individual behavior and health conditions, to intervene by sending information to the individual via devices such as their smartphone in order to change a behavior (Katapally et al., 2018). Smart devices can be deployed to monitor behaviour in teams to improve performance (Pentland, 2012), or monitoring student engagement to improve learning outcomes (Crosby & Ikehara, 2015).

The recent advances in Artificial Intelligence (AI) that we are currently experiencing—e.g., autonomous vehicles, facial recognition—have accelerated due to the combined developments of big data and analytics, especially machine learning. However, the origins of AI, and concerns over its adoption in public policy and administration, are much deeper. The early promise of AI in public sector practice centred on providing decision support for public managers (e.g., Barth & Arnold, 1999; Hadden, 1986; Hurley & Wallace, 1986; Jahoda, 1986; Masuch & LaPotin, 1989) but failed to materialize in any meaningful way. While the early promises of AI went unfulfilled, there have been dramatic advances in AI in recent years (Russell & Norvig, 2009) that could have important consequence for public management and governance. A key contributing factor to increasing maturity of AI technologies and the viability of AI application to public policy and administration is the increased availability of data that can be used to further machine learning. As algorithms become more widely used, increasingly autonomous, and invisible to those affected by their decisions, their status as impartial public servants becomes more difficult to monitor for bias or discrimination (Janssen & Kuk, 2016b). Today, AI systems are being used to detect irregularities, with aims such as reducing fraud and errors in service processing (Maciejewski, 2017). An even more speculative example (Death, 2015) addresses challenges of watershed governance, envisaging the application of AI to the continuous monitoring of complex streamflow dynamics and water chemistry and quality as part of decision support systems for communities concerned with environmental flows as well as crucial water supply. The possible extension to Artificial Intelligence that could offer, autonomously, better decisions than the community might make in resolving the conflicts around the vital tradeoffs among the many interests, human and ecological—as well, perhaps, as the rights of the river itself—is a topic of ongoing debate. Relatedly, the question of meaningful human involvement in decisions related to problems of human security has been addressed in a recent report on the role of AI in nuclear war (Geist & Lohn, 2018).

4. Discussion

Given the scan of examples of policy analytics in practice, where does this revision to the policy analysis model fit in the modern governance toolkit, and what do the examples of successful policy analytics applications tell us about the possibilities for its future, and the limitations it will likely face?

We must be careful not to overstate what policy analytics can tell us. Take, for example, the rhetoric around predictive analytics (Gandomi & Haider, 2015), which can serve as an input into framing a policy problem before it is apprehended as such. In 'predictive policing', where potential crimes, offenders, and victims are identified *a priori*, police resources can be directed proactively (Brayne, 2017; Perry, McInnis, Price, Smith, & Hollywood, 2013). The inherent complexities of social, economic, and behavioural phenomena, however, make policy prediction essentially impossible (Sawyer, 2005). While modeling for purposes of forecasting (Sims, 1986) and related approaches such as backcasting (Robinson, 1982) can serve as useful tools in policy analysis, and these techniques have improved with the increase in available data and growth in analytical capacity (Einav & Levin, 2014b; Wang, Kung, & Byrd, 2018), there are obvious limits on our ability to predict the future. Predictive models are necessarily abstractions from reality, and cannot feasibly include all individual and system factors. More likely are qualitative statements (including probability statements as to likelihood) about the direction of predicted change, including indications about possible unexpected outcomes. These are useful for policy analysis, especially in highly uncertain environments where unlikely events may still yield catastrophic outcomes.

It should be obvious that the proposed policy analytic approach will not solve all policy challenges. Despite the power of modern digital technology, a number of limitations and caveats remain. While more, and more accurate, evidence can improve our understanding and form the basis for better policy, we should not conflate the volume of big data with its representativeness. Despite the mesh of sensors that act as the collection net for policy-relevant data, there is the risk that those without the right devices or engaged in the targeted behavior may be rendered "digitally invisible" in the movement towards rapid policy design (Longo et al., 2017, p. 76). There are also a number of technical limits to assembling robust big data sets including challenges in data acquisition (especially where much of the really valuable data is closely guarded by private companies; (Golder & Macy, 2014; Verhulst & Young, 2018), data interoperability problems (Miller & Tucker, 2014), and legitimate privacy protections that place prohibitions on the sharing of data outside of programs or departments, or even on combining datasets behind protective firewalls. Even if data coverage is comprehensive, big data hubris can produce policy errors (Lazer, Kennedy, King, & Vespignani, 2014). Traditional social science designs research instruments to collect data in order to test a hypothesis, whereas big data analytics seeks to identify relationships (Wigan & Clarke, 2013). And the risk of apophenia—the seeing of patterns in random data—can lead policymakers to identify correlations that are easily mistaken for causal relationships (Boyd & Crawford, 2012).

Shi et al. (2017, p. 552) note that "only a few government decisions have already benefited from the systematic use of masses of data and evidence, and of cutting-edge modelling", with the norm being to rely on traditional forms of policy analysis. Several challenges are noted, centring on the democratic underpinnings of policy analysis. Since public sector problems typically involve making decisions on behalf of society at large, involving the allocation of public resources (Lerner & Lasswell, 1951), policy analytics must balance the need for robust analysis with the need to satisfy legitimacy expectations, transparency requirements, and opportunities for citizen participation.

Thus the policy analytics model—of the rapid prototype based on a digitally enabled system of communication, feedback, analytics and tool modification—does not apply across a wide range of policy problems or domains. Many policy areas are not amenable to minor policy tool modifications that can be communicated digitally. Few policy systems form such a tight linkage between a minor modification of a policy signal and an immediately detectable response from the system under observation, instead operating across long timescales between policy intervention and system response. Policy analytics is well suited to the digital realm of approaches such as A/B testing of government citizen service websites (Longo, 2018), whereas many policy decisions entail actions that have significant consequences diffused over many sectors. More often than not, the policy environment will be complex beyond the capabilities of even the most advanced analytics. The possibility of policy experimentation will apply in a limited set of circumstances, especially where legitimate ethical concerns could be raised.

Consider the 4-quadrant diagram in Figure 1, with the horizontal axis running from micro or local scale on the left through regional or meso-scale to global scale on the right, and the vertical axis from certainty as to system structure and environment at the bottom to profound uncertainty at the top. In the top right quadrant (high uncertainty, global scale) one has 'wicked problems', 'messes', concerns of post-normal science facing all the challenges of affective conflict and democratic dissent. Examples might be climate change, global hydrological cycle, poverty and inequality. But even in these challenging settings one can look to rapidly increasing computational capacity to develop decision support systems. To the extent that agreement can be achieved on appropriate policy objectives and instruments, there can be realistic ambitions for real-time policy analytic systems.

In the top left quadrant, more inclusive community engagement and deliberation, building on increasingly sophisticated decision-support systems, is feasible, but again expectations of integrated data analytic/policy analytic systems running on a real-time basis must rest on hopes for inclusive and collaborative policy formation processes building agreement on legitimate and acceptable policy objectives and norms of implementation. The lower right quadrant might be thought largely empty for the moment: there appear to be few global scale challenges for which one can have reasonable certainty around system structure and environment, except perhaps international agreements on classification systems or the like. But even here, as international agreements grow in number and specificity, policy analytic methods for monitoring and certifying compliance are increasingly significant.

Nevertheless, it seems that it is in the lower left quadrant, with reasonable certainty around the nature and context of micro or local scale problems that big data, data analytics and policy analytics can best support ongoing experimentation, continuous learning, policy formation, and adaptive management with effective implementation, monitoring and enforcement. Focusing on this quadrant, how might its boundaries evolve and expand? Evidently the operational problems faced in managing the direct provision of services at local level are more amenable to such experimentally-based adaptive control and self-regulation than for the problems that have to be addressed through cooperative federalism or similar institutional arrangements for negotiation among authoritative political units at different scales. Although the professional effort to differentiate the 'policy design' product from the more traditional language calls

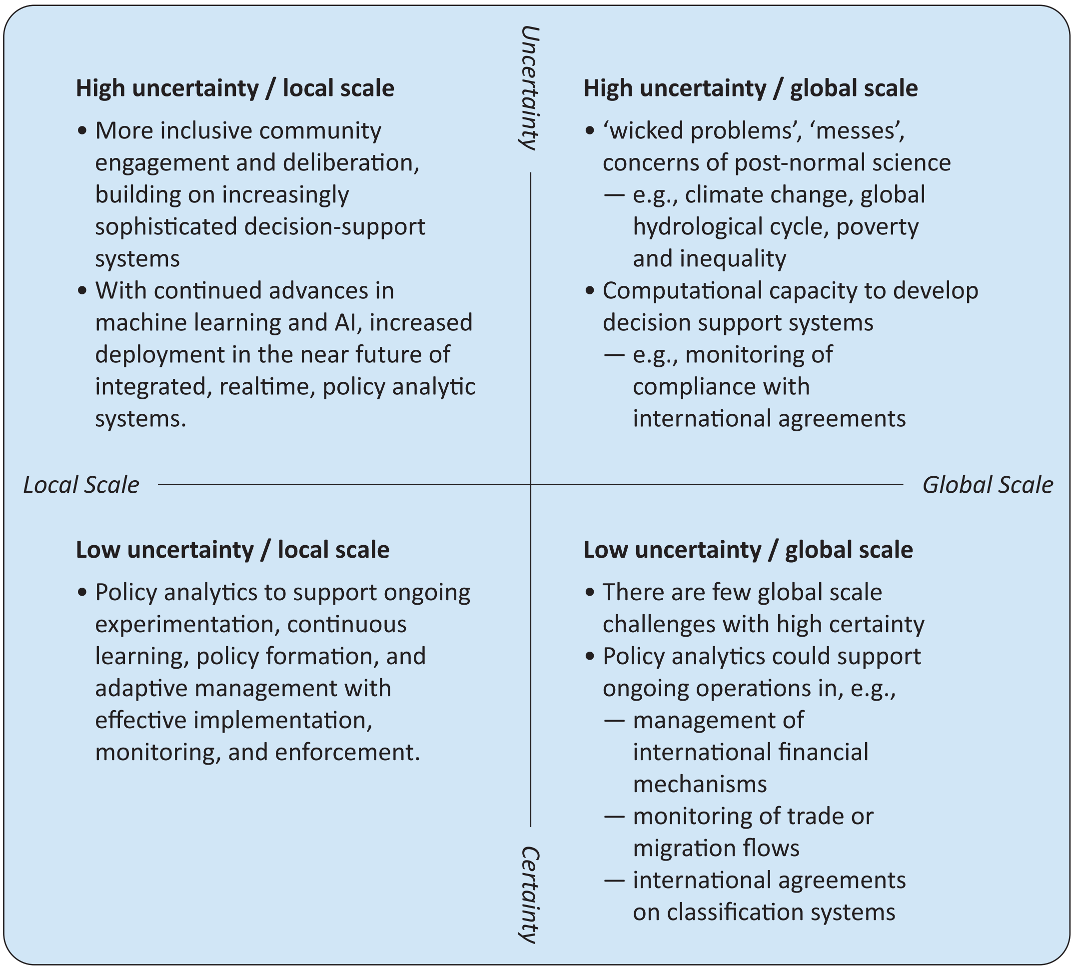


Figure 1. The applicability of policy analytics across scale and complexity.

attention more to the implementation end of the cycle than to the formulation portion, the bigger challenge for the rapid adaptation of design in response to user experience lies in the varied and slow instruments for implementing change in the operations of representative government, legitimately and with ongoing accountability. The fuzzy boundaries that separate a summative evaluation cycle for Cabinets or executive authorities from a formative evaluation cycle for management exercising delegated authority in decisions at small (how small?) scale might suggests limits to policy analytics—but they also suggest the potential of machine learning and autonomous and intelligent systems in pushing those boundaries far outward. The science fiction aspects of Joe AI, analyst, or Jane AI, authoritative decision-maker—and the challenges of teaching her/it in new schools of public policy—may be with us much sooner than expected, with consequent rapid advance in the spread of policy analytics as integrated system.

5. Conclusion

This article began with a quotation from a leading business magazine in 1973 that enthused about the possibilities of a policy supertool that then appeared imminent. That quotation was cited in a commentary from the Honourable C. M. Drury (then President of the Treasury Board of Canada—the agency charged with the development of tools for policy analysis and decision support in the Government of Canada at the time) in the inaugural issue of the journal *Canadian Public Policy*. In reaction to the fantastic possibilities envisaged, the Minister suggested that "While we may all have our occasional doubts about the advice offered by our traditional public servants, I am certainly not yet ready to trade them in on the strength of this promise!" (Drury, 1975, p. 91).

Almost a half-century later, does policy analytics represent the delayed realisation of that promised policy supertool, or yet another misplaced enthusiasm? Daniell et al. (2016, p. 11) conclude their special issue of policy analytics in practice with the consideration "that analytics have been somehow oversold", that political decision making can be overcome by masses of data, and deep analytics, producing automated solutions to any public problem. While evidence is important, decision making still requires judgment. New initiatives can be informed by past experience, but still require careful experimentation to avoid large implementation failures.

The emerging examples may be persuasive in their particular domains. But many of the problems confronted by policy analysts are indeed wicked problems involving differing time scales in complex systems where the effects of policy interventions on decisions and behaviour are unclear, uncertain, and of unknown duration. Much more crucially, agreement on the objectives or purposes of policy is usually lacking, and interests around the nature or instruments of policy intervention are conflicted. Not all policy environments are compatible with the policy analytics model. Much work remains to be done before we find the proper place for this promising development in an increasingly post-positivist, post-fact, post-truth world.

Acknowledgements

The comments of two anonymous reviewers, and the managing and academic editors for this thematic issue, are greatly appreciated.

Conflict of Interests

The authors declare no conflict of interests.

References

Alfaro, C., Cano-Montero, J., Gómez, J., Moguerza, J. M., & Ortega, F. (2016). A multi-stage method for content classification and opinion mining on weblog comments. *Annals of Operations Research*, *236*(1), 197–213.

Amy, D. J. (1984). Towards a post-positivist policy analysis. *Policy Studies Journal: The Journal of the Policy Studies Organization*, *13*(1), 207–211.

Aringhieri, R., Carello, G., & Morale, D. (2016). Supporting decision making to improve the performance of an Italian emergency medical service. *Annals of Operations Research*, *236*(1), 131–148.

Barth, T. J., & Arnold, E. (1999). Artificial intelligence and administrative discretion: Implications for public administration. *American Review of Public Administration*, *29*(4), 332–351.

Behavioural Insights Team. (2012). *Applying behavioural insights to reduce fraud, error and debt*. London: Cabinet Office. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60539/BIT_FraudErrorDebt_accessible.pdf

Blumsack, S., & Fernandez, A. (2012). Ready or not, here comes the smart grid! *Energy*, *37*(1), 61–68.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*(7415), 295–298.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication and Society*, *15*(5), 662–679.

Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review*, *82*(5), 977–1008.

Breckon, J. (2015). *Better public services through experimental government*. London: Alliance for Useful Evidence. Retrieved from http://www.capire.org/capireinforma/scaffale/Final.pdf

Business Week. (1973, May 19). Optimal control: A mathematical supertool. *Business Week*.

Cavallo, A., & Rigobon, R. (2016). The billion prices project: Using online prices for measurement and research. *The Journal of Economic Perspectives: A Journal of the American Economic Association*, *30*(2), 151–178.

Chan, J., & Bennett Moses, L. (2015). Is big data challenging criminology? *Theoretical Criminology*, *20*(1), 21–39.

Charalabidis, Y., Loukis, E. N., Androutsopoulou, A., Karkaletsis, V., & Triantafillou, A. (2014). Passive crowdsourcing in government using social media. *Transforming Government: People, Process and Policy*, *8*(2), 283–308.

Choi, H., & Varian, H. (2012). Predicting the present with Google trends. *The Economic Record*, *88*, 2–9.

Crosby, M. E., & Ikehara, C. S. (2015). Feedback from physiological sensors in the classroom. In T. Hammond, S. Valentine, A. Adler, & M. Payton (Eds.), *The impact of pen and touch technology on education* (pp. 381–387). Cham: Springer International Publishing.

Dahlhaus, P., Murphy, A., MacLeod, A., Thompson, H., McKenna, K., & Ollerenshaw, A. (2016). Making the invisible visible: The impact of federating groundwater data in Victoria, Australia. *Journal of Hydroinformatics*, *18*(2), 238–255.

Daniell, K. A., Morton, A., & Insua, D. R. (2016). Policy analysis and policy analytics. *Annals of Operations Research*, *236*(1), 1–13.

Death, R. G. (2015). An environmental crisis: Science has failed; let us send in the machines. *Wiley Interdisciplinary Reviews: Water*, *2*(6), 595–600.

Deleon, P. (1994). Reinventing the policy sciences: Three steps back to the future. *Policy Sciences*, *27*(1), 77–95.

De Marchi, G., Lucertini, G., & Tsoukiàs, A. (2016). From evidence-based policy making to policy analytics. *Annals of Operations Research*, *236*(1), 15–38.

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, *3*, 1376.

Drury, C. M. (1975). Quantitative analysis and public policy making. *Canadian Public Policy. Analyse de Politiques*, *1*(1), 89–96.

Dryzek, J. S. (1994). *Discursive democracy: Politics, policy, and political science*. Cambridge: Cambridge University Press.

Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2006). New public management is dead—Long live digital-era governance. *Journal of Public Administration Research and Theory*, *16*(3), 467–494.

Dutcher, J. (2014, September 3). What is big data?

datascience@berkeley. Retrieved from https://datascience.berkeley.edu/what-is-big-data

Einav, L., & Levin, J. (2014a). Economics in the age of big data. *Science*, *346*(6210), 1243089.

Einav, L., & Levin, J. (2014b). The data revolution and economic analysis. *Innovation Policy and the Economy*, *14*, 1–24.

Entwistle, V. (2003). Editorial. *Health Expectations*, *6*(2), 93–96.

Esperanza, L., & Dirk, P. (Eds.). (2014). *Making innovation policy work: Learning from experimentation*. Paris: OECD Publishing.

Fischer, F. (1995). *Evaluating public policy*. Chicago: Nelson-Hall.

Fredriksson, C., Mubarak, F., Tuohimaa, M., & Zhan, M. (2017). Big data in the public sector: A systematic literature review. *Scandinavian Journal of Public Administration*, *21*(3), 39–62.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144.

Geist, E., & Lohn, A. J. (2018). How might artificial intelligence affect the risk of nuclear war? *RAND Corporation*. Retrieved from https://www.rand.org/pubs/perspectives/PE296.html

Giest, S. (2017). Big data for policymaking: Fad or fast-track? *Policy Sciences*, *50*(3), 367–382.

Gerland, P. (2015). *Migration, mobility, and big data: An overview*. New York, NY: United Nations.

Golder, S. A., & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, *40*, 129–152.

Grubmüller, V., Götsch, K., & Krieger, B. (2013). Social media analytics for future oriented policy making. *European Journal of Futures Research*, *1*(1), 20.

Hadden, S. G. (1986). Intelligent advisory systems for managing and disseminating information. *Public Administration Review*, *46*, 572–578.

Harford, T. (2014). Big data: A big mistake? *Significance. Statistics Making Sense*, *11*(5), 14–19.

Harsin, J. (2015). Regimes of post-truth, post politics, and attention economies. *Communication, Culture and Critique*, *8*(2), 327–333.

Hartmann, H., Schuldt, B., Sanders, T. G. M., Macinnis-Ng, C., Boehmer, H. J.S., Allen, C. D., . . . Anderegg, W. R. L. (2018). Monitoring global tree mortality patterns and trends. Report from the VW symposium 'Crossing scales and disciplines to identify global trends of tree mortality as indicators of forest health'. *New Phytologist*, *217*, 984–987.

Höchtl, J., Parycek, P., & 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.

Hondula, D. M., Kuras, E. R., Longo, J., & Johnston, E. W. (2017). Toward precision governance: Infusing data into public management of environmental hazards. *Public Management Review*, 1–20.

Hurley, M. W., & Wallace, W. A. (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.

Jahoda, M. (1986). Artificial intelligence: An outsider's perspective. *Science & Public Policy*, *13*(6), 333–340.

Janssen, M., & Kuk, G. (2016a). Big and open linked data (BOLD) in research, policy, and practice. *Journal of Organizational Computing and Electronic Commerce*, *26*(1/2), 3–13.

Janssen, M., & Kuk, G. (2016b). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*, *33*(3), 371–377.

Jarmin, R. S., & O'Hara, A. B. (2016). Big data and the transformation of public policy analysis. *Journal of Policy Analysis and Management*, *35*(4), 715–721.

Javed, B., McClatchey, R., Khan, Z., & Shamdasani, J. (2018). A Provenance framework for policy analytics in smart cities. *arXiv [cs.CY]*. Retrieved from http://arxiv.org/abs/1804.07141

Johnston, E. W. (2015). *Governance in the information era: Theory and practice of policy informatics*. New York, NY: Routledge.

Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. *IT Professional*, *15*(6), 43–48.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kang, J. S., Kuznetsova, P., Luca, M., & Choi, Y. (2013). Where not to eat? Improving public policy by predicting hygiene inspections using online reviews. In *proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 1443–1448). Stroudsburg, PA: aclweb.org.

Katapally, T. R., Bhawra, J., Leatherdale, S. T., Ferguson, L., Longo, J., Rainham, D., . . . Osgood, N. (2018). The SMART study, a mobile health and citizen science methodological platform for active living surveillance, integrated knowledge translation, and policy interventions: Longitudinal Study. *JMIR Public Health and Surveillance*, *4*(1), doi: 10.2196/publichealth.8953

Khoury, M. J., & Ioannidis, J. P. A. (2014). Medicine. Big data meets public health. *Science*, *346*(6213), 1054–1055.

Kim, G.-H., Trimi, S., & Chung, J.-H. (2014). Big-data Applications in the Government Sector. *Communications of the ACM*, *57*(3), 78–85.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*(1), 1–14.

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication and Society*, *20*(1), 14–29.

Koene, A. (2017). Algorithmic bias: Addressing growing concerns. *IEEE Technology and Society Magazine*, *36*(2), 31–32.

Kumar, A., Nguyen, V. A., & Teo, K. M. (2016). Commuter cycling policy in Singapore: A farecard data analytics based approach. *Annals of Operations Research*, *236*(1), 57–73.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google flu: Traps in big data analysis. *Science*, *343*(6176), 1203–1205.

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., . . . Van Alstyne, M. (2009). Life in the network: the coming age of computational social science. *Science*, *323*(5915), 721.

Lee, D., Kim, M., & Lee, J. (2016). Adoption of green electricity policies: Investigating the role of environmental attitudes via big data-driven search-queries. *Energy Policy*, *90*, 187–201.

Lerner, D., & Lasswell, H. D. (1951). *The policy sciences. Recent developments in scope and method*. Stanford, CA: Stanford University Press.

Lindblom, C. E. (1959). The science of 'muddling through'. *Public Administration Review*, *19*(2), 79–88.

Longo, J. (2017). The evolution of citizen and stakeholder engagement in Canada, from Spicer to #hashtags. *Canadian Public Administration*, *60*(4), 517–537.

Longo, J. (2018). Digital tools for rapid policy design. In M. Howlett & I. Mukherjee (Eds.), *Handbook of policy design*. New York, NY: Routledge.

Longo, J., Kuras, E., Smith, H., Hondula, D. M., & Johnston, E. (2017). Technology use, exposure to natural hazards, and being digitally invisible: Implications for policy analytics. *Policy & Internet*, *9*(1), 76–108.

Longo, J., & McNutt, K. (2018). From policy analysis to policy analytics. In L. Dobuzinskis & M. Howlett (Eds.), *Policy analysis in Canada*. Bristol: Policy Press.

Lu, W., Chen, X., Ho, D. C. W., & Wang, H. (2016). Analysis of the construction waste management performance in Hong Kong: The public and private sectors compared using big data. *Journal of Cleaner Production*, *112*, 521–531.

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.

Majone, G. (1989). *Evidence, argument, and persuasion in the policy process*. New Haven, CT: Yale University Press.

Maragoudakis, M., Loukis, E., & Charalabidis, Y. (2011). A review of opinion mining methods for analyzing citizens' contributions in public policy debate. In E. Tambouris, A. Macintosh, & H. de Bruijn (Eds.), *Electronic participation (ePart 2011)* (pp. 298–313). Berlin: Springer.

Masuch, M., & LaPotin, P. (1989). Beyond garbage cans: An AI model of organizational choice. *Administrative Science Quarterly*, *34*(1), 38–67.

May, P. J. (1992). Policy learning and failure. *Journal of Public Policy*, *12*(4), 331–354.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review*, *90*(10), 60–68.

Meltsner, A. J. (1976). *Policy analysts in the bureaucracy*. Berkeley and Los Angeles: University of California Press.

Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big data in public affairs. *Public Administration Review*, *76*(6), 928–937.

Miller, A. R., & Tucker, C. (2014). Health information exchange, system size and information silos. *Journal of Health Economics*, *33*, 28–42.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2). doi:10.1177/2053951716679679

Monafu, D., Chan, S., & Turnbull, S. (2018). Is experimentation just the latest buzzword? *Policy Options*. Retrieved from http://policyoptions.irpp.org/magazines/february-2018/is-experimentation-just-the-latest-policy-buzzword

Morçöl, G. (2001). Positivist beliefs among policy professionals: An empirical investigation. *Policy Sciences*, *34*(3/4), 381–401.

Morgan, M. G., Henrion, M., & Small, M. (1992). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge: Cambridge University Press.

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *JAMA: The Journal of the American Medical Association*, *309*(13), 1351–1352.

Murphy, C. (2016). *Shared mobility and the transformation of public transit* (Research Analysis Report #TCRP J-11/TASK 21). Washington, DC: American Public Transportation Association. Retrieved from https://trid.trb.org/view.aspx?id=1401765

Narayanan, A., & Shmatikov, V. (2008). Robust deanonymization of large sparse datasets. In *IEEE symposium on security and privacy, 2008* (*SP 2008*, pp. 111–125). New York, NY: IEEE.

Newsham, G. R., & Bowker, B. G. (2010). The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: A review. *Energy Policy*, *38*(7), 3289–3296.

Noveck, B. S. (2018). Legal and other barriers to trust and innovation. In *Workshop on trust and opening governance*. Washington, DC: Georgetown University. Retrieved from http://opening-governance.org

Paris, C., & Wan, S. (2011). Listening to the community: Social media monitoring tasks for improving government services. In *CHI '11 extended abstracts on human factors in computing systems* (pp. 2095–2100). New York, NY: ACM.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Cambridge, MA: Harvard University Press.

Pentland, A. (2012). The new science of building great teams. *Harvard Business Review*, *90*(4), 60–69.

Perry, W. L., McInnis, B., Price, C. C., Smith, S., & Hollywood, J. S. (2013). *Predictive policing: The role* of crime forecasting in law enforcement operations. RAND Corporation. doi:10.7249/RR233

Persily, N. (2017). Can democracy survive the internet? *Journal of Democracy*, *28*(2), 63–76.

Pirog, M. A. (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.

Poel, M., Meyer, E. T., & Schroeder, R. (2018). Big data for policymaking: Great expectations, but with limited progress? *Policy & Internet*, *10*(3), 347–367.

Prpić, J., Taeihagh, A., & Melton, J. (2015). The fundamentals of policy crowdsourcing. *Policy & Internet*, *7*(3), 340–361.

Quade, E. S. (1975). *Analysis for public decisions*. New York, NY: Elsevier.

Quade, E. S. (1980). Pitfalls in formulation and modeling. In G. Majone & E. S. Quade (Eds.), *Pitfalls of analysis* (Vol. 8, pp. 23–43). Chichester: John Wiley & Sons.

Radin, B. A. (2000). *Beyond Machiavelli: Policy analysis comes of age*. Washington, DC: Georgetown University Press.

Raymer, J., Wiśniowski, A., Forster, J. J., Smith, P. W. F., & Bijak, J. (2013). Integrated modeling of European migration. *Journal of the American Statistical Association*, *108*(503), 801–819.

Robinson, J. B. (1982). Energy backcasting. A proposed method of policy analysis. *Energy Policy*, *10*(4), 337–344.

Russell, S. J., & Norvig, P. (2009). *Artificial intelligence: A modern approach* (3rd ed.). Essex: Prentice Hall.

Sawyer, K. R. (2005). *Social emergence: Societies as complex systems*. Cambridge: Cambridge University Press.

Schintler, L. A., & Kulkarni, R. (2014). Big data for policy analysis: The good, the bad, and the ugly. *The Review of Policy Research*, *31*(4), 343–348.

Shi, J., Ai, X., & Cao, Z. (2017). Can big data improve public policy analysis? In *proceedings of the 18th annual international conference on digital government research* (pp. 552–561). New York, NY: ACM.

Shulock, N. (1999). The paradox of policy analysis: If it is not used, why do we produce so much of it? *Journal of Policy Analysis and Management*, *18*(2), 226–244.

Sims, C. A. (1986). Are forecasting models usable for policy analysis? *The Quarterly Review*, *10*(1), 2–16.

Stone, D. A. (1988). *Policy paradox and political reason*. Glenview, IL: Scott Foresman & Co.

Till, B. C., Longo, J., Dobell, A. R., & Driessen, P. F. (2014). Self-organizing maps for latent semantic analysis of free-form text in support of public policy analysis. *Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery*, *4*(1), 71–86.

Tingling, P., & Brydon, M. (2010). Is decision-based evidence making necessarily bad? *MIT Sloan Management Review; Cambridge*, *51*(4), 71–76.

Tsoukias, A., Montibeller, G., Lucertini, G., & Belton, V. (2013). Policy analytics: An agenda for research and practice. *EURO Journal on Decision Processes*, *1*(1/2), 115–134.

Tufekci, Z. (2014). Engineering the public: Big data, surveillance and computational politics. *First Monday*, *19*(7). doi:10.5210/fm.v19i7.4901

Van Oort, N., & Cats, O. (2015). Improving public transport decision making, planning and operations by using big data: Cases from Sweden and the Netherlands. In *18th IEEE international conference on intelligent transportation systems, Las Palmas, Spain, 15–18 September 2015*. New York, NY: IEEE.

Verhulst, S. G., & Young, A. (2018). The potential and practice of data collaboratives for migration. *Stanford Social Innovation Review*. Retrieved from https://ssir.org/articles/entry/the_potential_and_practice_of_data_collaboratives_for_migration

Vie, L. L., Griffith, K. N., Scheier, L. M., Lester, P. B., & Seligman, M. E. P. (2013). The person-event data environment: Leveraging big data for studies of psychological strengths in soldiers. *Frontiers in Psychology*, *4*, 934.

Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, *126*, 3–13.

Ward, J. S., & Barker, A. (2013). Undefined by data: A survey of big data definitions. *arXiv [cs.DB]*. Retrieved from http://arxiv.org/abs/1309.5821

Welch, E. W., Hinnant, C. C., & Moon, M. J. (2005). Linking citizen satisfaction with e-government and trust in government. *Journal of Public Administration Research and Theory*, *15*(3), 371–391.

Wigan, M. R., & Clarke, R. (2013). Big data's big unintended consequences. *Computer*, *46*(6), 46–53.

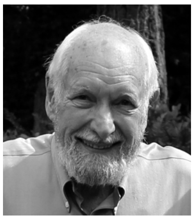
Wildavsky, A. (1979). Policy analysis is what information systems are not. In A. Wildavsky (Ed.), *The art and craft of policy analysis* (pp. 26–40). London: Palgrave Macmillan.

Xiong, G., Zhu, F., Dong, X., Fan, H., Hu, B., Kong, Q., . . . Teng, T. (2016). A kind of novel ITS based on space-air-ground big-data. *IEEE Intelligent Transportation Systems Magazine*, *8*(1), 10–22.

Zhou, K., & Yang, S. (2016). Understanding household energy consumption behavior: The contribution of energy big data analytics. *Renewable and Sustainable Energy Reviews*, *56*, 810–819.

About the Authors

Justin Longo is the Cisco Research Chair in Digital Governance and an Assistant Professor in the Johnson-Shoyama Graduate School of Public Policy at the University of Regina where he directs the Digital Governance Lab. He has a PhD in public policy and public administration from the University of Victoria (2013) and was a Postdoctoral Fellow in open governance at Arizona State University. His current research focuses on the organizational and societal implications of advancing technology.



Rod Dobell is Emeritus Professor of Public Policy at the University of Victoria, and Senior Research Associate of the Centre for Global Studies and the Centre for Public Sector Studies. He completed his PhD in economics at the Massachusetts Institute of Technology, and taught at Harvard University, the University of Toronto, and the University of Victoria. In 1991 he was named as the first appointee to the Francis G. Winspear Chair for Research in Public Policy at the University of Victoria.

COGITATIO

Politics and Governance (ISSN: 2183–2463)
2018, Volume 6, Issue 4, Pages 18–28
DOI: 10.17645/pag.v6i4.1586

Article

Realising the Benefits of Integrated Data for Local Policymaking: Rhetoric versus Reality

Hannah Durrant [1,*], Julie Barnett [1,2] and Emily Suzanne Rempel [1,2]

[1] Institute for Policy Research, University of Bath, Bath, BA2 7AY, UK; E-Mails: h.durrant@bath.ac.uk (H.D.); j.c.barnett@bath.ac.uk (J.B.), e.s.rempel@bath.ac.uk (E.S.R.)
[2] Department of Psychology, University of Bath, Bath, BA2 7AY, UK

* Corresponding author

Submitted: 24 May 2018 | Accepted: 17 October 2018 | Published: 21 November 2018

Abstract
This article presents findings from local government projects to realise the benefits of big data for policy. Through participatory action research with two local statutory authorities in the South West of England, we observed the activities of identifying, integrating and analysing multiple and diverse forms of data, including large administrative datasets, to generate insights on live policy priorities and inform decision-making. We reveal the significance of both data production and policymaking contexts in explaining how big data of this kind can be called upon and enacted in policy processes.

Keywords
big data; integrated data; local government; policymaking

Issue
This article is part of the issue "Big Data Applications in Governance and Policy", edited by Sarah Giest (Leiden University, The Netherlands) and Reuben Ng (National University of Singapore, Singapore).

1. Introduction

The claims made for big data in business contexts are well established (e.g., Mayer-Schönberger & Cukier, 2013). Kitchin (2014b) discusses the powerful sets of discourses that are employed to support the application of big data to realise tangible improvements to business processes, products and profits. These include, but are not limited to, the ability of big data technologies to enhance logistics planning, reduce inefficiencies, understand customer preferences, target products and services to new and existing markets and combat fraud (Kitchin, 2014b, pp. 117–123). More recently, attention has turned to the potential of big data for policymaking settings (e.g., United Kingdom [UK] Parliament, 2015), and the challenges involved in harnessing this potential to realise policy aims and objectives for the public good (Janssen, Konopnicki, Snowdon, & Adegboyega, 2017; Kennedy, Moss, Birchall, & Moshonas, 2015; Malomo & Sena, 2016; Schintler & Kulkarni, 2014). Administrative (open) data is particularly prominent in Poel, Meyer and Schroeder's (2018) analysis of the use of big data in policymaking, being used in two thirds of the 58 such initiatives they identified. Questions have been raised about how and where in the 'policy cycle' big data-derived analysis could feed in (Höchtl, Parycek, & Schöllhammer, 2015), with increasing emphasis being placed on the role that data can play in predicting need and defining policy priorities for the future (Giest, 2017; Malomo & Sena, 2016). This work usefully disaggregates the applications of data, moves beyond rhetoric and opens up thinking about the spaces for data science to inform policymaking.

However, policymaking processes are not straightforward or linear, and there is a need to theorise the social contexts of both data production and policymaking to understand the boundaries and barriers to big data for policy in practice. We set out to reveal the temporally-specific and contingent ways in which data are articulated in the demand for evidence, and discuss how the practices and preoccupations of policymaking both shape and are being shaped by the promise of data.

The article unfolds as follows. We begin (in section two) by rehearsing the claims that have been made about big data, and that have sought to give this ubiquitous but simultaneously elusive term some definitional clarity. We focus on claims made about the promise of data for policymaking, and problematize assumptions of linear and rational policymaking processes into and through which data science can flow. We rather propose a counter theory of policymaking as struggles over the right to advance ideas about policy; why it is needed, what it should do, for whom, how and to what end (Carmel & Papadopoulos, 2003). We argue that it is in these deeply value-laden and political contexts that data are produced and repurposed, and insights are allowed, or otherwise, to be admitted as a form of evidence.

Section three briefly describes the participatory action research approach adopted in this project and details the partnership and processes by which the project progressed. Section four presents findings and reflections from the project; focusing on the ways in which data is constituted as relevant to policymaking, the terms on which its use is resisted; and the importance of relationships of trust to underpin data processes in practice. We conclude in section five by discussing the significance of the social context of both data generation and policymaking to explain what can actually be done with data in policy settings.

2. Big Data and Policy Making

Conventional attempts to define big data have tended to focus primarily on its characteristics; initially emphasising its volume, variety, and velocity (see Kitchin & McArdle, 2016). A more recent proliferation of characteristics identified with big data (e.g., Uprichard, 2013) has rendered the term more, rather than less, opaque (Kitchin & McArdle, 2016). In an attempt to isolate the most salient qualities of big data, Kitchin (2014b) stresses the distinction between small data sources—based on a population sample, infrequently collected and processed slowly and periodically—and big data that is both exhaustive (n = all) and generated and reported in close to real time. For Kitchin and McArdle (2016) the two most important characteristics of big data sources are exhaustivity and velocity.

Consideration of the sources of data has also been used to ground understanding of what is commonly considered to constitute big data (Connelly, Playford, Gayle, & Dibben, 2016). Data are being generated from a greater variety of sources than ever before (Kitchin, 2014a). Some of this data is what Mayer-Schönberger & Cukier (2013, p. 113) refer to as ‘data exhaust’; the by-product of people’s digital activities and interactions (e.g., financial transactions and social media activity), repurposed to another end. For Connelly et al. (2016), the ‘found’ nature of big data, and its ability to be valuably repurposed, is a significant feature. They differentiate between data that is ‘made’ by social scientists to study the social world, and data generated for entirely different purposes yet possessing considerable research utility (see also Cowls & Schroeder, 2015).

The identification of a wide array of characteristics and sources of big data conveys a sense of its ubiquity, but also the extent to which it has defied definitional clarity. Recent scholarship has begun to systematise ‘types’ of big data according to the types of traits that it possesses (Connelly et al., 2016; Kitchin & McArdle, 2016). Both of these articles identify multiple types or forms of big data. In particular, Connelly et al. (2016) assess the extent to which administrative data, a form of data derived in the process of administering services and systems and commonly held by UK government (nationally and locally) and other public sector bodies, can be considered one form of big data. They argue that it meets the conditions because it is often exhaustive, highly granular, large (as a consequence of being both exhaustive and highly granular), messy and unstructured and, importantly, found and repurposed rather than made (see also Kitchin & McArdle, 2016; Malomo & Sena, 2016).

We would agree with this assessment. In our experience, working with statutory bodies in the South West of England, we found that local government administrative data, particularly when integrated with other forms of demographic, contextual and unstructured data, demonstrated many of the characteristic of big data. Data relating to, for instance, primary and secondary health care, social benefit claims, and the delivery of public services, cover the entire population (i.e., all patients, claimants and service recipients within an administrative boundary). In addition, administrative data are often produced in real time and can be extracted for use at frequencies close to real time. They are granular to the extent that they are individual-level and contain extensive fields; including details of service use, as well as demographic, service process and background information. Granularity is enhanced further where data is linked and integrated, and we found that some datasets contained both structured and unstructured data (e.g., case notes and service user comments and feedback). Most importantly, however, these data were found to be of value to social science research and policymaking, rather than made.

Furthermore, and in line with other scholars that have focused on the benefits of big data for policymaking, we include administrative data as a source of big data—particularly where it is linked and integrated with other data sources—on the grounds of its particular relevance and value for policy (Connelly et al., 2016; Poel et al., 2018). Administrative records provide governments at all levels in the UK with unique access to diverse data generated on the people and communities they serve, and there is a growing literature on the application of these kinds of data in policy settings (Janssen et al., 2017; Malomo & Sena, 2016; Poel et al., 2018).

Current data initiatives are accompanied by powerful rhetoric about the significance of big data for policy, emanating from within policymaking communities. In 2015, the UK Parliament identified harnessing the benefits of big data as a key issue for government(s); describing data as “the new oil” (UK Parliament, 2015) and, just as is the case in business contexts, here too the claims for the possibilities afforded by big data are expansive. Stephan Shakespeare, in his review of Public Sector Information, enthusiastically asserts that “from data we will get the cure for cancer as well as better hospitals; schools that adapt to children’s needs making them happier and smarter; better policing and safer homes; and of course jobs” (Shakespeare, 2013, p. 5). Thus, as well as implying a set of characteristics, the term ‘Big Data’—coined in the context of a data revolution that government(s) in the UK are keen to capitalise on—is also pervaded by a set of strongly held and asserted beliefs about the purposes to which data can be put and the ends that are envisaged (Kitchin, 2014b; Markus & Topi, 2015). Markus and Topi (2015) contend that definitions should acknowledge big data as more than sets of data with particular characteristics that require novel analytical techniques, and equally recognise the ideas that seek to inspire its use (see also boyd & Crawford, 2012). They argue for viewing big data as, “a cluster or assemblage of data-related ideas, resources and practices” (Markus & Topi, 2015, p. 3).

Optimistic claims for the potential of big data tend to obscure challenges associated with its use. At the most extreme, big data advocates promote a view that data—in great enough volume and when properly interrogated—can “speak for themselves” (Anderson, 2008), and that this may be a welcome step forward for evidence-based policymaking (Mayer-Schönberger & Cukier, 2013). More recently, more critical approaches have questioned the portrayal of data as neutral and data science as objective; raising the politics of data capture and analysis.

A growing critical data studies literature (Iliadis & Russo, 2016) has emphasised that data is generated, curated, processed and interpreted through frameworks that determine what is constituted as data and how it can be translated into information (boyd & Crawford, 2012; Kitchin, 2014a). Such frameworks are inherently political because of what they count and what they leave out, what they make visible and what they render invisible; particularly when being visible and counted is a necessary precondition in qualifying for political, economic and social resources. As Johnson (2015) states: “the ability to make one’s group, and one’s interests legible to the state, organizations, or other individuals is increasingly determined by where one stands in the data”.

Kennedy et al. (2015, p. 175) point to a widespread awareness—particularly among social scientists—of the ways in which data is shaped and given value by the context in which it is produced and the methods by which it is aligned, processed and analysed. Following them, we sought to understand the extent to which data, and the techniques for extracting meaning from it, came under critical examination in the practices and processes of policymaking. We note that the claims regarding the potential of big data for policymaking are often disconnected from the sets of ideas, resources and practices involved in data application to policy. This article is concerned with understanding the narratives, processes and practices by which data can meaningfully grease the wheels of decision-making in policy settings.

Recent scholarship has sought to identify opportunities for big data insights to inform policymaking, by focusing on the stages of the policy cycle most amenable to injections of data-derived evidence. Höchtl et al. (2015) journey through the steps involved in policy making—e.g., agenda setting and discussion, policy formation and decision-making, implementation, etc.—providing reflections on the potential contribution of big data to each. They particularly highlight the possibility for real-time data processing to enable continuous evaluation throughout the process. Giest (2017) explores government use of a range of administrative and real-time data to design and customise policies. She highlights the value of these data to agenda setting and policy implementation. Malomo and Sena (2016) describe a case study of using integrated data in local government and highlight the benefits of big data for predicting need and effectively targeting services.

The studies usefully break down and compartmentalise the different functions of big data for policy making—options appraisal, predictive analysis, real-time evaluation etc. However, they tend to overplay the extent to which policymaking proceeds stepwise, through a series of linear stages, and understate the challenges associated with the straightforward inflow of any kind of information and evidence (Cairney & Heikkila, 2014).

Rather than seeing policymaking as a linear and rational process, we start from the premise that policymaking is the variable outcome of consensus, negotiation, contestation or co-option of ideas about what is to be done, by whom, how and for what purpose (Carmel & Papadopoulos, 2003). Ideas embodied in narratives of causation compete for the right to be accepted. Power and context influence the strength of the narrative to succeed (Jessop, 2009; Stone, 1989). Policymaking is a messy process in which conflicting ideas and policies are brought forward, debated, and implemented.

Scholarship is emerging on how data and data technologies fit into a narrative-conflict view of policymaking. Kettl (2016) emphasises that the nonlinear nature of policymaking problematises the assumption that data is used simply as evidence to make the best policy choice (see also Poel et al., 2018). They argue that good data analysis is useless without a good narrative. In contrast, Janssen and Helbig (in press) argue that data technologies have great potential to interrupt the status quo and revolutionise policymaking.

In summary then, the ideas that foreground government(s) enthusiasm for realising the potential of big, integrated forms of data have primarily focused attention on the potential of technical innovations. However, processes of data science in policy settings are embedded in dynamic, multifaceted, and deeply political contexts of problem definition, evidence interpretation, solution identification and decision-making. These settings materially affect the ways in which big data is called upon and able to impact decision-making. We engaged with local government activity around integrated data in order to consider how data informs policymaking processes: how the practices and preoccupations of the policy process define and shape the generation and use of data science; and how integrated data, as one form of evidence generation, shapes and redefines these policy practices.

3. Methods

This article presents a series of observations drawn from participatory action research within a set of local government data projects that ran at different times and for different durations between 2013 and 2018. Together, these projects set out to realise the benefits of integrated administrative and other data to policy development and practice at the local level, with the ultimate aim of establishing, testing and evaluating processes to change the culture of data use within and across public services.

Given the project aim, the approach was grounded in the principles of participatory action research (PAR) (Bergold & Thomas, 2012; Coghlan & Brydon-Miller, 2014) in a cycle of data collection and analysis, reflection and action, that emphasised equal collaboration between researchers and practitioners, trust and discretion in communication and the production of shared knowledge (Brydon-Miller, Greenwood, & Maguire, 2003). This approach was applied to four contemporary policy priorities for the statutory authorities involved (see Table 1). Working within the tenets of PAR, the research conducted within these four settings utilised data linked and anonymised by the statutory authorities prior to release for project purposes, and sought to contribute to both the development of data-informed policy and practice, and wider understanding of the contexts, processes and practices that realise the benefits of data for local government. In this article we present our observations from these projects.

The core project team that worked across all four policy priorities included three researchers from the University of Bath Institute for Policy Research (the authors) and three senior policy officials from two local statutory bodies within the South West of England. In the course of the projects the team engaged with service managers from relevant departments, and with other policymaking bodies and civil society organisations in the region. These included commissioning managers with responsibility for setting policy priorities; business analytics officers and managers from commissioned services and voluntary sector delivery bodies. The size and composition of the wider stakeholder group involved varied considerably between projects, a point relevant to understanding variation in the conditions under which data-derived evidence can inform policy and practice and which we reflect on in the findings section below. The project activities were instantiated within a formal collaboration agreement between the three core institutions, which detailed data management and use protocols, and received ethical approval from the University of Bath.

Table 1 outlines the four settings for the research and the associated data sources used to inform decision-making.

In each case, the projects progressed through discussion with the project team and wider stakeholders to understand the policy issues and context; define policy questions of interest; identify and access potential sources of data; conduct and interpret quantitative analyses (e.g., propensity score matching, cluster analysis and predictive analysis). Insights from the analysis often raised additional questions, and policy questions were refined, and additional data and analysis sought accordingly. This process of making sense of the data and deciding next steps took place within regular fortnightly meetings of the project team at the University, as well as ad hoc meetings with other policy actors involved in each of the settings when each project was 'live'. In addition to the comprehensive notes taken of all of these meetings our reflections and observations draw on email exchanges, telephone conversations and the content of and comments on project documents (including, for example, project scoping documents and reports of the analyses).

4. Findings and Reflections

In drawing together the projects and seeking to explore the interactions between the policy context, policy questions and data integration practices, we present findings and reflections under three themes. Firstly, we consider the way in which the *relevance* of data is constituted in policy settings, as a function of its perceived value in answering policy questions. Secondly, we explore the conditions under which data applications to policy are *resisted*. Finally, we reveal significant aspects of the *relationships* between different interested parties where data and policymaking intersect.

4.1. Relevance of Data to Address Policy Questions

In using integrated data in local government settings, policy questions, not data, were the starting point for data projects. Whether the issue was financial hardship, designing health and wellbeing services or education service provision, it was the policy questions and context that defined the scope for data to inform decision-making. In this context, data did not "speak for themselves" (Anderson, 2008). Its potential utility to policymaking was realised where it was deemed able to be relevant to, and admitted (along with other evidence) as a

Table 1. Policy priorities, aims and data sources.

Policy Priorities	Policy Aims	Indicative Data Sources
Financial hardship	To understand the consequences of economic downturn and austerity for financial wellbeing.	*Individual-level:* • Time-series: social benefit claims, employment status, household composition, disability status • Demographic information *Aggregate-level (Lower Super Output Area):* • Debt (County Court Judgements) • Household composition • Social benefits claims • Tax credits • Income deprivation
Community health services	Review community health services for patients with a particular chronic condition to understand the efficacy of these services and the effect on health outcomes.	*Individual-level:* • Time-series: secondary care records, in-patient admissions, out-patient appointments, co-morbidities and clinical test results; • Attendance at community care services. • Demographic information *Aggregate-level (GP Surgery):* • Patient population; • Health checks. *Additional data collected:* • Patient illness perceptions and experience of services.
Wellbeing services	Review and redesign community wellbeing services.	*Individual-level:* • Wellbeing service administrative records: participant numbers, dates and service location; • Case notes; • Evaluation and outcomes. *Additional data collected:* • Interest in wellbeing services; • Various measures of personal wellbeing; • Demographic information; • Provider experiences of delivering wellbeing services.
Education services	Understand changed profiles of demand and redesign education services.	*Individual-level:* • Time-series: school and academies census; • Pupil demographic information; • Educational needs and status; • Free school meals eligibility.

response to a policy enquiry. In other words, the value, or otherwise, of data was constituted only in the context of critical examination of what data could represent and what it could say—given how it was generated, curated, processed and integrated—and only in relation to policy questions. Even where there were large volumes of exhaustive data, the application of that data to local policymaking was contingent on what data was considered able to illuminate about the perceived problem and what was allowed to be asked about it. The weight of data, typically associated with the big data phenomenon, did not unproblematically transfer into a weight of evidence (Schintler & Kulkarni, 2014).

Having said that, the projects do illustrate how a keen interest in the power of data, particularly the potential of combining multiple forms of disparate data, is reinvigorating and reshaping the demand for evidence in policymaking processes at the local level. Policy partners were keen to identify and explore the benefits of the vast amounts of data routinely collected to inform service development, and were, in some cases, open to broadening the options for policy change in light of the subsequent insights.

There was sometimes an absence of data deemed sufficiently relevant to addressing particular policy questions. As an example we discuss the case of the review of local health services, which explored patient pathways and outcomes through services relating to a particular condition. In a routine appraisal of these services policy officials were interested in understanding barriers to and enablers of service take-up. They had a clear view about the nature of the policy problem: low levels of service take-up among certain patient groups in particular areas—and a set of questions predicated on assumptions about policy options for service improvement. However, project discussions with the research team led them to broaden their enquiries. They commissioned a Rapid Evidence Assessment (REA) to extend their understanding of the factors influencing service uptake. A REA is an evidence synthesis that follows a systematic methodology but, in order to be rapid, is restricted in breadth, depth and comprehensiveness compared to a systematic review (Barends, Rousseau, & Briner, 2017). The REA raised explanations for low service take up and variation in service performance that were not previously part of the scope of the data project. This called into question the sufficiency of the data that had previously been designated as relevant to informing the policy question.

Policy makers became aware that data routinely collected and available on these services largely served to facilitate service administration and audit (e.g., by providing information on volume of provision, attendance and dates) rather than understanding reasons for low service take up review performance. They recognised gaps in the data relating to patient experience, as well as patient health management behaviours. In this case they decided to collect additional survey data. The survey drew together a number of existing validated scales (including the Illness Perception Questionnaire) and the sample size was all patients. The responses were combined at the level of the individual with existing administrative data to inform their decision-making.

In contrast, in other projects, the boundaries of the policy issue were broader and questions more loosely specified. For example, the enquiry into the consequences of economic downturn and austerity began with the broad aim to utilise linked data to identify changes in frequency and intensity of financial hardship at the local level. Equally, the review of education services began with a general aspiration to better understand changes in the profile of demand. In these cases, formulation of the policy questions and defining and deciding on the scope for data enquiries progressed through a series of incremental, iterative steps. Here, policymaking tended to be in response to emerging policy issues where there were numerous stakeholders advancing competing narratives about the nature of the problems and seeking to shape the range of acceptable policy responses. Thus unlike the healthcare case above, here the framing of the policy questions and legitimate solutions were contested. Despite policy officials' enthusiasm to realise the potential of integrated data, broadly defined questions raised challenges for identifying the types of data that could usefully provide answers. In the education service case, policy officials and service managers initially struggled to conceptualise how the various data on pupils and schools that they held could be exploited. The breadth of policy questions rendered the sources of relevant data that could address the questions as opaque.

In these cases, seeking to establish the existence and/or the relevance of data often involved conversations between the core project team and other data holders—often service managers in departments within the two local statutory authorities but outside the area of direct policy interest. This then involved a second-stage of iteration, to establish the validity of the data access request and legitimise the relevance of the data. In the health and wellbeing and the understanding financial hardship case studies, access to data held by other service providers was denied on the grounds that the resource cost of providing data was greater than the perceived benefit to policy. Combining data involves multiple sites where judgements are made about the relevance of data to policy questions that may not be owned or of interest to those that hold the data.

Issues of data relevance are also circumscribed by the divisions of local and national policy responsibilities. In the case of the data enquiry into the impact of economic downturn and austerity, the insights drawn from an analysis of combined datasets on levels of benefit claiming, employment status, county court judgements, household composition, physical health and other factors, showed particular groups of people (in work on low pay) as potentially more exposed to financial hardship. However, the ability of policy officials to action this insight was restricted, as it was deemed outside the scope of local policy. This case illustrated that insights from available and relevant data may not be actionable. This may be for a range of reasons—in this case, local government action was precluded by national government ownership of what transpired to be the issue where action was required.

4.2. Resistance to Data Use in Policy

The projects provided examples of ways in which the application of data to inform policy was challenged and resisted. For example, policy officials disputed or sought to discredit the legitimacy of data use where they had reservations about its quality. Sometimes claims about poor data quality were substantiated with reference to the purposes for which it had been generated: reservations were expressed around the notion that data collected for one reason should be repurposed for another. On other occasions resistance was focused on the way in which the dataset had been constructed where reservations focused on the validity of repurposing particular variables. Anticipation of public perceptions about the re-use of data also served to bolster concerns and augment resistance to data use.

In all of the projects, concern was raised about the potential impact on re-appropriation of the data of missing observations, human error and biases resulting from how they were collected, maintained and stored. In the financial hardship case study, policy officials resisted the inclusion of certain data fields on the grounds that the values they contained may be incorrect. For example, they questioned the quality of some demographic information in one data set where individual characteristics had not been crucial to determining service eligibility. Similarly, in the wellbeing services case, data related to the provision and uptake of these services (e.g., numbers of participants) were perceived to be more systematically collected—and thus more accurate—than evaluation data or data on participants' health outcomes. It was the evaluation and health outcome data, however, that was of greater value and significance in the re-appropriation of the data and the potential for linking with other data sets. Thus in both these examples, the extent to which data was considered suitable for reuse was related to the social context in which the data had originally been compiled: the likely motivations underlying the inclusion of particular variables and imputations about the care with which the data set had been constructed.

Further challenges to the validity of data applications for policy were raised in the education services case. Here the legitimacy of repurposing the data was less about the accuracy of the data and more about the validity of extrapolating from it. The example of data on eligibility for free school meals (FSM) illustrates this point. Even where data was perceived to be recorded correctly (i.e., all eligible registrations for FSM were input on data systems), policy officials highlighted that the introduction of universal infant Free School Meals in 2014 had significantly affected the numbers of parents registering their child's eligibility (Sellen & Huda, 2018). The perceived effect of this policy change was that FSM data had lost its value as an indicator of changed profiles of demand for education services.

In all of the cases, it was not that policy officials lacked curiosity and enthusiasm for harnessing the value of existing data. Indeed, aspirational ideas circulating within and beyond local government (e.g., Mayer-Schönberger & Cukier, 2013; Shakespeare, 2013) about the vast potential of big data permeated their thinking and motivated their efforts to realise the benefits for policymaking. However, the processes of data curation highlighted that the ability to be curious was tempered by the contexts in which datasets were compiled, structured and maintained in local government settings. For example, it was clear in the financial hardship case that a consequence of decisions to hold personal data on clients only for the time that they were service users was that datasets tended to over-represent continuous, and longer-term service users, thus obscuring patterns in short-term and cyclical service use.

To some extent the limitations inherent to data collection and management terms were perceived by policy officials to be a consequence of data protection compliance; specifically the requirements—under the Data Protection Act 1998 (Information Commissioner's Office, n.d.-a) and the (at the time forthcoming) General Data Protection Regulation and Data Protection Act 2018 (Information Commissioner's Office, n.d.-b)—to only collect and retain as much personal data as is necessary, and not to reuse data in ways incompatible with the original purpose. Where there were limits on data applications given the terms under which data had been generated, policy officials were reluctant to revisit consent and tended to opt for the narrowest interpretation of their ability to generate or reuse data. This thus limits "extensibility" (Mayer-Schönberger & Cukier, 2013, p. 109), whereby the ability of data to have multiple uses is intentionally embedded in data collation protocols.

In addition, even where legal compliance was assured, policy officials were often juggling between two competing narratives about public perceptions of data use by local government. While they recognised a sense of public expectation that they would use available data 'smartly' to innovate and better target services, in practice they were also stifled by anticipation of public reservations about the acceptability of linked data. In other words, in their use of data policymakers recognised a distinction between what is legally defensible and what may be considered ethically permissible.

As a consequence, emerging awareness of data to answer policy questions did not unproblematically translate into availability of data. Policy makers' sensitivity to data quality and legitimacy, the legality of its use and the anticipated responses of the public could lead to data being rendered inadmissible in integrated data projects. Professional tacit knowledge was used to ground data, counteract its inaccuracies, navigate its ethical and legal implications and mitigate the likelihood of misreading the insights that it can yield. Data was only admissible where policy professionals could first fill in blanks and inaccuracies with their local knowledge of how things *actually* are.

4.3. Relationships with Data and Policy

This final section presents significant aspects of the relationships that effect the intersections between data and policymaking. We first observe that trust is vital to enable integrated data projects to have value in policy settings and then consider how the politics of policymaking impacted data sharing and the terms of engagement for different stakeholders.

Throughout the project collaboration, data was sourced and released in stages as trust in the partnership—between members of the core project team and the wider stakeholders—was built over time, ethical and legal boundaries established and the value of early analyses realised. For example, in the community health services case, establishing the policy-research relationship led to the project partners first seeing the potential value of conducting a RER, and then being confident to act on the relationship this showed between patient perception of illness and health management behaviours by collecting attitudinal data that could be linked with secondary health care records.

The data projects proceeded via an abductive approach—flip flopping between patterns emerging in the data and hypotheses, seeking additional insights and testing further hypotheses. For instance, in the example above, having refined the initial scope of the enquiry in the light of the RER, mini hypotheses to explain low service take up by certain patient groups were proposed, tested, discussed and revised in relation to the policy context. Across each of the projects, the rationale for additional data releases was grounded in the cementing of trust in the partnership and the realisation of benefits from the preceding stages. Thus, the value of the collaborative data enterprise was realised through processes that iteratively established confidence in the partnership.

Sometimes relationships between the project partners were more problematically embedded in the politics of data sharing; for example between levels and departments of government, between different public services, and between the policy partners and the research team. Some data—for example individual-level data on unemployment and take-up of employment services—were held nationally by the Department for Work and Pensions and unavailable to local policy officials on the grounds that it would breach their terms for information governance. Thus relevant data on variance in financial wellbeing was only available to the project in aggregate form.

On one occasion in the community health services project, difficulties in obtaining data from a service provider were attributed to the politics of the commissioner-provider relationship between the statutory authority and the provider. Given the nature of this relationship—and the unequal power relations within it—the senior policy officials within the core project team reflected that the other party may have been unwilling to share data for fear that the data would be misappropriated beyond the scope of the project and used to monitor their performance. This speaks to the significance of trust and transparency over purpose as well as methods in integrated data projects. Concern about the potential for data to surveil service performance was particularly apparent where ideas about policies—what they intend to achieve, for whom and how—were disputed. For example, in the wellbeing services project, service providers were unwilling to share data with service commissioners where they felt exposed when sharing data showing low volumes of activity without taking into account the quality of provision for vulnerable clients. A further variation on this theme was observed in the review of education services. Here data analysis was sought by service managers where it gave confidence to pursue preferred explanations for changed profiles of demand. Alternative explanations were undermined by questioning data accuracy or by citing particular aspects of policy context.

A final example from the financial hardship case, of the importance of trust was evident in a debate between one of the policy partners and a third sector organisation. The dispute centred on the scale of financial hardship in the local area and the nature of services required in response. Third sector providers made reference to a range of evidence to support their position. Significantly, the data held by these third sector providers was not made available for integration as they claimed that its collection was conditional on particular sets of expectations for use. Their contention was that the data had been shared with them precisely because they were distinct from local government and a source of support for those wishing to raise grievances about local government. As a result they considered that sharing these data with local authorities would be a breach of trust. This provides a further illustration of how limits on linking data are not restricted to technical issues about the availability or format of data—rather they are shaped by relational considerations around trust and the politics of data and policymaking.

5. Discussion and Conclusions

The findings and reflections from our project to realise the benefits of data for policy have revealed particular sets of ideas about data (Markus & Topi, 2015). These concern the ways in which the relevance of data is socially constituted in policy settings and the conditions under which data applications to policymaking can be and are resisted, as well as the degree to which the relationships between stakeholders at the intersection of data and policy influence what data processes and insights can be considered. Overall, we highlight that variation in the degree to which integrated data and the techniques of data science are able to encroach on policy practice, is contingent on the ideas about and social context and processes of both data generation and policymaking.

The ambition to utilise the vast quantities of data that local government produces and can access is driven, at least in part, by the motivation to realise the aspirational claims made about big data for policymaking. However, the projects we draw on highlight the first-and-foremost requirement to be problem-oriented in big data applications to policy. Even where we observe the seeming ubiquity of data, there are still circumstances where we have data for which there aren't questions and questions for which we do not have data (boyd & Crawford, 2012; Kennedy et al., 2015); and it is questions and not data that drive policy calls on evidence.

In contrast to early definitions of big data that focused on the characteristics of data (volume, variety, velocity) with less reference to the purposes to which it could or should be put, we find that where integrated data is applied to policymaking its most defining quality is its ability to be big in value (Cowls & Schroeder, 2015; Organisation for Economic Co-operation and Development [OECD], 2013). In policy settings the value of data is allied to its ability to provide insight germane to live and pertinent policy and practice preoccupations. We find that the choice of what data to use or collect involves problem-based decisions on what would be indicative of the thing(s) we are trying to understand.

Given this grounding for the potential of data for policy, the social contexts and processes involved in data generation, maintenance and storage become of vital importance. It is these contexts and processes that determine what data can, and what it cannot, represent and say. We have shown that administrative datasets tend to function primarily as a tool to audit public services; telling us how many services are delivered, for how many people and when. As such, their reuse value is limited where the aim of data applications to policy enquiry is the curious exploration of social phenomena, to understand what could work better, for whom and under what conditions.

The value of integrated data to policy challenges is further exacerbated when consideration is given to the errors and biases data contains as a consequence of how it is arrived at; what priorities are ascribed to its accuracy; and what legitimacy and legality it has when it is repurposed. The implication of these considerations is that the existence of large quantities of data is not an asset in itself to local policymaking. Its value can only be realised if and when the constraints of the social contexts and processes of its production can be mitigated. Even then, we have shown that the potential value of data is conditional on the political context in which policy is being made.

We have shown considerable differences in the contexts in which local policies are made. These contexts are not fixed and static, but highly variable, multifaceted and contingent on the historical trajectory of policymaking in the field. The context shapes ways of acknowledging problems and justifying the solutions to which policy is aimed.

Policymaking takes place on different timescales depending on the mode of policymaking. For instance, whether policymaking is happening as part of a routine programme of on-going review, or in response to an unanticipated shock—such as a public (media) outcry, a change in national or regional policy, or a change in social/economic circumstances—that disrupts routine policymaking processes and 'normal' policy timetables. At any given time, policy concerns can accelerate up through the rankings of priorities, or become suddenly subordinate to other more pressing preoccupations.

Big data analytics, with its focus on quick, novel and exploratory enquiry (Höchtl et al., 2015; Mayer-Schönberger & Cukier, 2013), could be seen to align well with extraordinary and fleet of foot policymaking; often seen as happening at a pace that traditional methods of information generation can't match (Whitty, 2015). However, such an assessment of the potential impact of big data-derived evidence underplays the complexity and politics of policymaking, particularly at points of disruption—for example times of economic downturn and austerity. In our experience, both times of routine policy appraisal and urgent reaction to policy crisis involve, first and foremost, the advancement and debate of *ideas about policy*, as well as related *ideas about data* (Markus & Topi, 2015) and what constitutes evidence.

The extent to which policy problems and potential options are tightly defined and agreed upon differs in different policy context. Ideas about policy, data and evidence are contained within a political reality that shapes and delimits the boundaries of policy aims; the purpose to which it can be addressed, the extent to which ownership and responsibility over the domain is open or closed, and the degree of disagreement and dispute among stakeholders over the aims and purpose of policy. The nature of the policymaking context and the issues being explored affects what questions can legitimately be asked of big data and the ways in which the resultant insights are considered as admissible as evidence that can form the basis for decision-making. Issues vary in the degree to which they are contested, how urgent they are, how open, how risky, etc. As a consequence, we find that in practice highly contested local welfare policy has a qualitatively different profile of considerations shaping the 'pull' on data science than, for example, the temporarily more consensual context of local health service provision for patients with a particular chronic condition.

Thus in our exploration of how the practices of data intersect with the practices and preoccupations of policy, we find a more nuanced and politically contingent call on data than would be suggested by the rhetoric around the potential of data. Indeed, we suggest that rather than looking at data science as a technical aspect of government activity underpinned by expansive claims for the power of data, we should instead see data science as contingent on the ideas, realities and political contexts of government practice. Scholarship and practice around these topics must be alert to both the potential impact of data on policymaking but also the ways in which the practices of making policy condition the potential for data to be used.

Acknowledgments

We are grateful to all the policy officials who allowed us to be part of the local government integrated data projects.

Conflict of Interests

The authors declare no conflict of interests.

References

Anderson, C. (2008, June 23). The end of theory: Will the data deluge make the scientific method obsolete? *Wired*. Retrieved from https://www.wired.com/2008/06/pb-theory

Barends, E., Rousseau, D. M., & Briner, R. B. (Eds.). (2017). *CEBMa Guideline for rapid evidence assessments in management and organizations (Version 1.0)*. Amsterdam: Center for Evidence Based Management.

Bergold, J., & Thomas, S. (2012). Participatory research methods: A methodological approach in motion. *Forum: Qualitative Social Research*, *13*(1), 1–31. doi:10.17169/fqs-13.1.1801

boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, *15*(5), 662–679. doi:10.1080/1369118X.2012.678878

Brydon-Miller, M., Greenwood, D., & Maguire, P. (2003). Why action research? *Action Research*, *1*(1), 9–28. doi:10.1177/14767503030011002

Cairney, P., & Heikkila, T. (2014). A comparison of theories of the policy process. In P. Sabatier & C. Weible (Eds.), *Theories of the policy process* (3rd ed.). Chicago, IL: Westview Press.

Carmel, E., & Papadopoulos, T. (2003). The new governance of social security in Britain. In J. Millar (Ed.), *Understanding social security* (pp. 31–52). Bristol: Policy Press.

Coghlan, D., & Brydon-Miller, M. (Eds.). (2014). *The SAGE encyclopedia of action research* (Vols. 1–2). London: SAGE Publications Ltd.

Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, *59*(C), 1–12. doi:10.1016/j.ssresearch.2016.04.015

Cowls, J., & Schroeder, R. (2015). Causation, correlation, and big data in social science research. *Policy and Internet*, *7*(4), 447–472. doi:10.1002/poi3.100

Giest, S. (2017). Big data for policymaking: Fad or fast-track? *Policy Sciences*, *50*(3), 367–382. doi:10.1007/s11077-017-9293-1

Höchtl, J., Parycek, P., & Schöllhammer, R. (2015). 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. doi:10.1080/10919392.2015.1125187

Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, *3*(2), 1–7. doi:10.1177/2053951716674238

Information Commissioner's Office. (n.d.-a). Guide to data protection. *Information Commissioner's Office*. Retrieved from https://ico.org.uk/for-organisations/guide-to-data-protection

Information Commissioner's Office. (n.d.-b). Guide to the general data protection regulation (GDPR). *Information Commissioner's Office*. Retrieved from https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr

Janssen, M., & Helbig, N. (in press). Innovating and changing the policy-cycle: Policy-makers be prepared! *Government Information Quarterly*. doi:10.1016/j.giq.2015.11.009

Janssen, M., Konopnicki, D., Snowdon, J. L., & Adegboyega, A. (2017). Driving public sector innovation using big and open linked data (BOLD). *Information Systems Frontiers*, *19*(2), 189–195. doi:10.1007/s10796-017-9746-2

Jessop, B. (2009). Cultural political economy and critical policy studies. *Critical Policy Studies*, *3*(3), 336–356. doi:10.1080/19460171003619741

Johnson, J. A. (2015). What is the data in big data? *Discover Society*. Retrieved from http://discoversociety.org/2015/08/03/on-the-frontline-what-is-the-data-in-big-data

Kennedy, H., Moss, G., Birchall, C., & Moshonas, S. (2015). Balancing the potential and problems of digital methods through action research: Methodological reflections. *Information, Communication & Society*, *18*(2), 172–186. doi:10.1080/1369118X.2014.946434

Kettl, D. F. (2016). Making data speak: Lessons for using numbers for solving public policy puzzles. *Governance*, *29*(4), 573–579. doi:10.1111/gove.12211

Kitchin, R. (2014a). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 1–12. doi:10.1177/2053951714528481

Kitchin, R. (2014b). *The data revolution: Big data, open data, data infrastructures & their consequences*. London: Sage Publications Ltd.

Kitchin, R., & McArdle, G. (2016). What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1), 1–10. doi:10.1177/2053951716631130

Malomo, F., & Sena, V. (2016). Data intelligence for local government? Assessing the benefits and barriers to use of big data in the public sector. *Policy and Internet*, *9*(1), 7–27. doi:10.1002/poi3.141

Markus, M. L., & Topi, H. (2015). *Big data, big decisions for science, society, and business*. Waltham: Bentley University. Retrieved from https://www.bentley.edu/files/2015/10/08/BigDataWorkshopFinalReport.pdf

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data. A revolution that will transform how we live, work and think*. London: John Murray.

Organisation for Economic Co-operation and Development. (2013). *Exploring data-driven innovation as a new source of growth: Mapping the policy issues raised by 'big data'* (OECD Digital Economy Papers, No. 222). Paris: OECD Publishing. doi:10.1787/5k47zw3fcp43-en

Poel, M., Meyer, E. T., & Schroeder, R. (2018). Big data for policymaking: Great expectations, but with limited progress? *Policy and Internet*, *10*(3), 347–367. doi:10.1002/poi3.176

Schintler, L. A., & Kulkarni, R. (2014). Big data for policy analysis: The good, the bad, and the ugly. *Review of Policy Research*, *31*(4), 343–348. doi:10.1111/ropr.12079

Sellen, P., & Huda, N. (2018). Evaluation of universal infant free school meals. *Education Policy Institute*. Retrieved from https://epi.org.uk/publications-and-research/evaluation-universal-infant-free-school-meals

Shakespeare, S. (2013). *Shakespeare review: An independent review of public sector information*. London: Department for Business, Innovation and Skills. Retrieved from https://www.gov.uk/government/publications/shakespeare-review-of-public-sector-information

Stone, D. A. (1989). Causal stories and the formation of policy agendas. *Political Science Quarterly*, *104*(2), 281–300. doi:10.2307/2151585

United Kingdom Parliament. (2015). Making big data available: Key issues for the 2015 Parliament. *United Kingdom Parliament*. Retrieved from https://www.parliament.uk/business/publications/research/key-issues-parliament-2015/technology/big-data

Uprichard, E. (2013). Big data, little questions. *Discover Society*. Retrieved from https://discoversociety.org/2013/10/01/focus-big-data-little-questions

Whitty, C. (2015). What makes an academic paper useful for health policy? *BMC Medicine*, *13*(301), 1–5. doi:10.1186/s12916-015-0544-8

About the Authors

Hannah Durrant leads the research programme of the University of Bath Institute for Policy Research. Her research interests are the governance of economic and social policy reform and the role of data, evidence and expertise in policymaking. She is currently working on several projects related to big data in government.

Julie Barnett is a Professor of Health Psychology at University of Bath. Her research interests and projects span public and stakeholder understandings of risk, risk communication, management of food allergy and intolerance, loneliness and the way in which new forms of data are used as evidence in policy making.

Emily Rempel is a Researcher at the University of Bath in the Department of Psychology and the Institute for Policy Research. Her work focuses on the intersections between publics, policy, and data technology with an emphasis on feminist perspectives. Other research interests include health and well-being, epidemiology, and population mental health.

COGITATIO

Politics and Governance (ISSN: 2183–2463)
2018, Volume 6, Issue 4, Pages 29–38
DOI: 10.17645/pag.v6i4.1565

Article

Big Data under Obama and Trump: The Data-Fueled U.S. Presidency

Barbara Trish

Political Science Department, Grinnell College, 50112 Grinnell, USA; E-Mail: trish@grinnell.edu

Submitted: 2 May 2018 | Accepted: 28 September 2018 | Published: 21 November 2018

Abstract
The much-heralded use of data, analytics, and evidence-based decisions marks the U.S. presidency, wherein many processes and decisions are structured by the analysis of data. An approach with historical precedent, reliance on data was prominent under Obama, and is even under Trump, despite signals to the contrary. This article examines three cases from the Obama era: microtargeting in electoral campaigns, performance management in government, and signature drone strikes employed by the national security apparatus. It also reflects on the early Trump administration. The processes described are highly dependent on data, technically big data in two instances. The article examines the cases both on their own terms and in the context of a critical lens that directs attention to the political economy of the data. The analysis helps unpack the allure of data and analytics as well as the challenges in structuring an environment with a measured approach to data and big data, which would examine both their potential and drawbacks.

Keywords
analytics; big data; data; drone strikes; evidence-based; microtargeting; Obama; performance management; president; Trump

Issue
This article is part of the issue "Big Data Applications in Governance and Policy", edited by Sarah Giest (Leiden University, The Netherlands) and Reuben Ng (National University of Singapore, Singapore).

1. Introduction

"Evidence-based" is the 21st century coin of the realm, with broad, seemingly unbounded applicability. Practices relying on evidence, their counterpart "data-based decisions", along with the tracking, data, and analytics that fuel all have infused the public realm, private lives and all areas between. Presidential politics in the U.S. is no exception.

Reliance on data, in some cases technically "big data", marks the contemporary presidency to the extent that it has become the default approach, part of its institutional DNA. In some measure, this mirrors the broad progression of thought and practice that extends beyond the narrow scope of the U.S. presidency, from politics and governing generally, to commerce, sports, and all variety of enterprises that value efficiency, either as means of deploying resources or as an ultimate goal. But for U.S. presidents and their administrations, reliance on data has attendant advantages, quite apart from the operations and decisions that unfold—and these serve to deflect attention from problems, including difficult ethical challenges, that accompany the data-driven presidency. Because of this, the presidency needs the rare combination of expertise and detachment to yield effective decision making, in order to weigh the traps and biases associated with this world of data against the advantages.

This article focuses primarily on the role of data in the Obama administration, in structuring processes and decision-making, with cases drawn from three different domains—presidential selection, internal governance, and tactical national security decision-making. Early indications for the Trump presidency suggest consistency in the role of data, despite some signals to the contrary. Taken together, these cases demonstrate how reliance on data is buttressed by recurring calculations that emphasize efficiency, and in some cases a private sector that provides the data, at an extreme even extracting data from individuals without compensation. In other words, the political economy of data in the presidency helps explain its modern allure—and recognizing it can also inform prudent action in the future.

2. Epistemological and Political Roots

The turn to data in the U.S. federal executive is consistent with early twentieth century ideas, both epistemological and ideological. Emphasizing data has roots in logical positivism, which envisioned knowledge as the result of empirical evidence. In the U.S., the ideological counterpart to positivism was turn-of-the-century Progressivism, valuing information and expertise as tools to disrupt the stranglehold of moneyed interests and patronage-fueled political parties on political power (Greider, 1992; Orren & Skowronek, 2017). But as the century advanced, information and expertise in the domain of the executive became vehicles not for disruption, but for leadership by the president in the U.S. separated system, with a massively-expanded state apparatus. In other words, data and expertise became instrumentally valuable in politics.

Like the political world, the academy pivoted to embrace empirical evidence. About when Progressive political sentiment took hold, reformers in the American academy moved to establish a distinctive approach to the study of politics and government, shedding "the legalist and theoretical way in which political life was studied in the European academy" (Susser, 1992, p. 4). The result was a new discipline of political science, at times with some academics inserting themselves into the rough and tumble of politics. But the more pervasive quality of this new approach—which was fully formed by the 1960s—was its social-scientific orientation, emulating the scientific model and placing a premium on empirical, especially quantitative, evidence.

Against this backdrop, contemporary applications of data in the U.S. presidency—that is, reliance on evidence-driven practices and data—are not fundamentally new. Indeed, well before "analytics" and "big data" emerged, there was a strong element of data-based politics in the U.S., extending over the nation's entire history and with remarkable scope. From census data informing allocation of congressional representation starting at the founding, to data-fueled economic projections mandated by the New Deal, and even to Ronald Reagan-as-president using polling data to refine his rhetoric, the historical examples are abundant. However, politics and governing in the U.S. has reached a critical juncture, with reliance on data so pervasive that it's difficult to imagine anything but; it has become the default choice, the go-to solution for decisions, management and administration. With this in mind, the following section describes how data are employed in three different domains of the contemporary U.S. presidency, beginning with campaign politics, marked by what can justifiably be called big data.

3. Data and Big Data under Obama

3.1. The Data Science of Campaigns

While the road to the White House had been paved with data and analytics for some time, the near obsession of Barack Obama's two campaigns with evidence-based practices represents a difference in kind. What's more, the campaigns' successes impelled the wide diffusion of the data-centric campaign model. Even Donald Trump, who conveyed skepticism about data—at times eschewing it—subscribed to fundamentals of a data-driven campaign model.

The predominant narrative of both Obama's 2008 nomination and general election wins emphasized that the campaign's data-driven operation successfully mobilized voters, especially new ones, to the polls. In 2008, a ground game flush with money fueled a sophisticated data-rich field operation, enhanced by online capacity which included new platforms to engage voters. The 2012 addition to the narrative emphasized that the reelection campaign was metric-driven and informed by the insights of social scientific research. Both campaigns fundamentally ran on data, not unprecedented in approach, but certainly in scope.

The data at the heart of mobilization efforts are voter lists, used by campaigns to identify potential supporters and mobilize them—through direct contact—to the polls. These basic lists are longstanding, in fact the byproduct of the early Progressive Era introduction of voter registration. Ironically, the information collected by this turn-of-the-nineteenth-century reform, meant to weaken political parties, became the raw material for the mobilization efforts of the parties and their candidates. Before the advent of polling, these lists provided a rare portrait of relative party strength among the electorate as well as measures—like party affiliation and demographics collected by the state—that could inform mobilization efforts (Hersh, 2015, p. 49).

Obama's voter contact data were simply an advanced version of those early lists, but digitally enhanced and readily operational through a user interface. The campaign used "VoteBuilder", the Democratic Party's propriety data, accessed through a user interface purchased from a left-leaning for-profit, NGP-VAN. VoteBuilder lists offer—at their core—the same type of information that in a prior era a party agent might have retrieved from official voting records, namely voting history and demographics of the registered vote. Now these lists are augmented to include additional individual-level information about the voter, drawn from a number of sources, including commercial firms, as well as parties and campaigns themselves which glean information from field staff and volunteer interactions with voters.

By 2008, the use of data like these was common, not just for presidential campaigns. So too were forays in microtargeting, procedures to further augment data by means of statistical analyses, a process that had been evolving over the prior decade. Microtargeting techniques produce synthetic measures of voter characteristics—"model scores"—by means of predictive analytics, integrating the results of large-n survey data with the augmented voter file. The model scores serve as criteria for a particular voter contact effort, targeted to specific individuals. Indeed, the "micro" aspect of this enterprise involves data analysis at the *individual* level, with targeted contact efforts similarly aimed. It is, in effect, an algorithmic decision-making process, though the practice began well before the term was applied to it.

Obama in 2008 engaged in a virtually uninterrupted process of modeling and refining the data, and then modeling again (Issenberg, 2012a). Progressive data giant Catalist, another for-profit firm partnering with the Obama effort, extended the data beyond the traditional cache of registered voters to include unregistered voters, not represented in traditional voter files. The Catalist data offering, according to the firm's own accounts, numbered 265 million cases, reflecting approximately the universe of voting-age adults in the U.S.

Conventional wisdom holds that the prowess of Obama's resource-flush 2008 campaign contributed heavily to his win. But by 2012, the reelection efforts would be enhanced significantly by an evidence-based understanding of the effectiveness of voter mobilization techniques, drawing heavily from experimental research with ties to the social-scientific community. The 2012 campaign, gripped by a culture of experimentation, employed evidence-informed programs (IEPs) not only in direct voter contact protocols used in the field, but also in digital and fundraising campaigns (Issenberg, 2012b).

The 2012 campaign was structured to give data and analytics a strong voice. At the Chicago Obama headquarters, a team of fifty analytics professionals worked out of the "the cave", with a direct line of reporting to the Chief Innovation and Integration Officer, who reported directly to the campaign manager. Data and technology departments constituted an estimated 30–40% of headquarter staff, and "[a]nalytics was the breakout star of 2012" (Engage, 2012). The data-focused structures and practices were complemented by the attitudes and norms of personnel, including senior staff with a willingness "to listen to numbers people rather than consultants acting on old-fashioned political intuition" (TechPresident, as cited in Engage, 2012.)

These 2008 and 2012 campaigns represented state-of-the-art data operations—in fact "big data", at least in the sense that they integrated data from a variety of sources, augmenting them repeatedly with newly-acquired information. The data even approached an N = all (Mayer-Schöenberger & Cukier, 2013) quality, in Catalist's case approximating the universe of the voting age population. But not all data that fuel the presidency are big in this sense. In fact, data collected under management protocols are rather conventional, despite being part of a monumental data-collection enterprise.

3.2. Tracking the Executive with Performance Management

Performance management in the abstract focuses on how efficiently and effectively the executive branch administers the programs of the federal government. It's data-driven management, which—much like the use of data in campaign politics—has become systematic and elaborate over time, especially since the 1990s. The data at the heart of performance management as practiced in the Obama years were collected by agencies, permitting judgment of the extent to which the outcomes of their activities met the goals of the programs and of the administration.

Most observers trace the development of performance management to the Bill Clinton era, though the impetus to employ management to consolidate the president's leadership of the executive branch came earlier, during the Richard Nixon administration. Nixon, in an effort to harness the discretion of the federal agencies and to ensure agreement across the administration with his policy priorities, layered on management responsibilities to the existing budget office, creating the Office of Management and Budget (OMB) in 1971. While OMB offered an institutional arm to the president for management, it wasn't until later, in 1993, that performance management as a systematic approach was codified by Congress in the 1993 Government Performance and Results Act (GPRA), requiring federal agencies to engage in strategic planning every five years and undertake annual performance reviews. And then in 2010, Congress passed the GPRA Modernization Act (GPRAMA), which revised specific expectations for performance management, including movement from annual to quarterly reviews/reporting, making performance management an ongoing process. This world is data-heavy, requiring that agencies establish goals and track progress toward achieving them. The initial statute instructed agencies to develop "quantifiable and measurable program targets" as well as "outcome measures", metrics by which the real-world success of the programs would be judged (Harris, 2015, pp. 105–106).

In its rhetoric, the Obama administration embraced performance management. It built onto the efforts of the Bush administration (Jochum, 2009), which had itself prioritized performance management and had devised and touted its Performance Assessment Rating Tool (PART), a quantitative assessment of goals and performance used by over 200 federal programs, estimated to account for 20% of the federal budget. Jeff Zeints, the acting director of OMB at the start of the Obama administration, described GPRA and Bush's PART as important starting points for the new administration. Ratcheting up the hype, Obama's performance management was spearheaded by "performance guru" Shelley Metzenbaum, who was responsible for developing www.performance.gov, a tool to both articulate the administration's approach and to provide access to copious reports and reviews filed under the program.

A wide variety of data is collected under performance management protocols, with each agency establishing annual performance goals for its mission areas, identifying metrics for assessing the goals, and then reporting the actual performance. Much of the data is straightforward, the product of counting and tracking. For example, the U.S. Department of Agriculture (USDA) reports the number of wetland acres (in millions) restored by the Conservation Reserve Program (CRP), a measure compiled from CRP contracts. But other measures entail estimation procedures, requiring technical expertise and skill, with results that may be marked by irregularities. Consider USDA's metric for enabling access to healthcare facilities for rural areas. Measured as the percentage of people who are provided access to new and/or improved essential community facilities, this metric involves estimation of a geographic service area for each facility and the population within it, both difficult to assess (USDA, 2013). Messiness aside, this is not "big data" by any conventional definition, though monumental in scope, since every agency collects measures to assess a large number of goals. To get a sense of the magnitude, consider the 2013 performance management reports for Commerce and USDA alone, each listing over forty metrics, while Housing and Urban Development (HUD) reported more than fifty. Twenty-four agencies are subject to this data-driven management process.

Data collected by the General Accountability Office (GAO) offer a valuable window to the experiences of agency managers in performance management, and the 2013 Federal Managers Survey reveals a mixed picture about the implementation of performance management. Administered in late 2012 and early 2013, the survey sampled a population of some 148,000 managers and supervisors in agencies that undertake performance management as specified by the two statutes, producing approximately 3,000 usable responses. The survey found that a full 48.8% had never even heard of the GPRAMA. Just under one-fifth (18.8%) of managers suggested that performance data were not easily accessible to them, while close to one-third (31.2%) felt the data were not easily accessible to their employees. When asked whether data were formatted in a manner easy to use, 27.7% reported they were not, and 30.2% expressed that they did not have sufficient analytic tools to collect, analyze and use the data (GAO, 2013). These indicators can be taken as glass half-full or half-empty, but regardless suggest that the reality of data-based performance management under Obama might not have met the promise of the accompanying build-up.

That said, performance management is transparent, with not only the data publicly available, but also the perception of the practitioners revealed systematically. Furthermore, the data are quite traditional, metrics with which to judge performance and, presumably, instruct decision makers. The following example offers a dramatic contrast, drawn from the realm of national security, involving the specification of targets for military attack.

3.3. High-Stakes, Covert Signature Strikes

To base decisions about whom to attack on evidence is unsurprising. But the Obama administration employed a controversial technique to target individuals or groups of individuals for drone strikes when they bore the characteristics—the "signature"—of those likely to be engaged in terrorist activity. In contrast to "personality strikes", in which the U.S. targets known terrorists (Zenko, 2012a), signature strikes are based on patterns of behavior indicative of terrorists, even if the individual target is not known to be a terrorist. President George W. Bush was the first to authorize such signature strikes (Zenko, 2013, p. 12), though this came at the end of his tenure.

These strikes are shrouded in secrecy, extreme even compared to the typical opacity of national security. Micah Zenko notes that signature strikes have not been acknowledged officially. "[N]o U.S. Government official has ever acknowledged the practice of signature strikes". Nor has any official "described the practice, justified it, or explained how it is consistent with the…laws of war" (M. Zenko, personal communication, 26 April 2018). Even more so, information about precisely what data and analytical tools inform the targeting is sketchy, with what is known owing largely to Edward Snowden's June 2013 leaks of National Security Agency (NSA) data.

Snowden revealed that the NSA mines metadata, essentially the trail that follows digital and cell-phone communication (Hu, 2017, p. 235). Even absent the content of the communications, these metadata allow the analyst, most likely relying on a combination of machine learning and network analytical techniques, to identify potential terrorists by their patterns of connections to others, including to known terrorists. Journalist Glenn Greenwald emphasizes that the validity of the data is not confirmed by traditional techniques, like engaging "operatives or informants on the ground" (Scahill & Greenwald, 2014). This threat of imperfection is captured in the oft-repeated comment, attributed to an unnamed State Department official: "[T]he C.I.A. sees 'three guys doing jumping jacks'…[and] thinks it is a terrorist training camp, [adding that those] loading a truck with fertilizer could be bombmakers—but they might also be farmers" (Becker & Shane, 2012).

Making decisions based on incomplete evidence is not new to the world of military tactics. Zenko (2012b) recounts an anecdote conveyed by General Colin Powell about his early history in Vietnam:

> I recall a phrase we used in the field, MAM, for military-age male. If a [helicopter] spotted a peasant in black pajamas who looked remotely suspicious, a possible MAM, the pilot would circle and fire in front of him. If he moved, his movement was judged evidence of hostile intent, and the next burst was not in front but at him.

The pilot in Powell's account makes a judgment based on the available evidence, flawed as it might be. The drone strike likely reflects similar judgment—but with split-second processing, the application of algorithms to data, perhaps even real-time geo-location data. Notably missing in this endgame is direct human judgment. Hu (2017, p. 231) asserts that the absence of human judgment distinguishes this big-data approach from the "small-data" methods of the past, which relied on human perception and human decision making. The other significant element is that the algorithmic process yields a quantitative measure of likelihood that a person is a terrorist—or that a targeted geographic space would encompass terrorists.

Critics find especially concerning that the decision to kill is based on a likelihood generated by an algorithm. But perhaps more problematic is the high civilian death toll associated with drone strikes. Data from New America Foundation (NAF), Long War Journal (LWJ), and The Bureau of Investigative Journalism (TBIJ), mostly from 2004–2012, estimate over 400 drone strikes in Pakistan, Yemen and Somalia, with approximately 12% (401) citizens among the 3,430 killed (Zenko, 2013, p. 13.)

Advances in technology—drones, the widespread use of mobile technology, as well as the ability of the NSA and CIA to track and analyze the exhaust—have opened the door for signature strikes. And while the details of the data and analytics that undergird them escape public scrutiny, it's clear that this is an executive branch big-data enterprise, not just in terms of volume, but also in the substitution of machine judgment for human judgment.

4. *Plus Ça Change*…in Early Trump?

Donald Trump, despite being an unconventional candidate and president, over the two years of his administration has signaled that in many respsects he follows in the footsteps of his predecessor regarding the use of data. This orthodoxy, however, is especially notable given that the president has at times vocally eschewed evidence-based practices. Plenty of time remains for the Trump approach to data to take shape, but at this juncture it looks like rhetoric does not always mesh with actions.

Candidate Trump expressed disdain for campaign data, calling it "overrated" (Vogel & Samuelsohn, 2016b), but then assembled a rather conventional voter data and mobilization effort, admittedly smaller and flying under the radar more so than his predecessor's (Vogel & Samuelsohn, 2016a). By early 2016, the operation was staffed by two former Republican National Committee (RNC) operatives, low key in orientation, but with experience working with the RNC's Voter Vault, the counterpart to the Democratic National Committee's VoteBuilder, which had fueled Obama's and—eventually—Hillary Clinton's campaigns. Largely undetected, the Trump campaign assembled "Project Alamo", an ambitious digital database that aided in online and offline targeting, strategic decisions and voter mobilization—as well as a dose of voter *demobilization*, attempting to limit the Hillary Clinton vote (Green & Issenberg, 2016).

And then there was Cambridge Analytica. Trump turned to the U.K.-based data and analytics firm, which was later revealed to have misappropriated Facebook data for the purpose of its "psychographic modeling" activities. The New York Times, working with London's Observer, reported that Cambridge Analytica, with close ties to central figures in the Trump orbit like Steve Bannon and the Mercers, acquired personal information on Facebook users by means of an academic, who claimed the data were for the purpose of academic research (Rosenberg, Confessore, & Cadwalladr, 2018).

While the tangled web of the Facebook data breach and possible connection to Russian collusion remained unresolved by late-2018, finance reports confirm that Cambridge Analytica was a player in the Trump data operation. Federal Election Commission (FEC) records from 2016 compiled by the Center for Responsive Politics (CRP) show disbursements of $5.9 million from the Trump campaign to the data firm. Notably, these disbursements were dwarfed by the $87.8 million paid to Giles-Parscale, the San Antonio digital marketing firm that was responsible for Project Alamo.[1]

Trump's embrace of data in the campaign phase is replicated in management of the administration, and there's even sign of the same Obama-era promotional voice. However, the ends to which Trump's performance management are directed are distinctively-Trump: to limit the reach of the federal government. Of course, this same goal is advanced by the record-number of key appointed positions in the executive branch unfilled well into the term (Kruzel, 2018) and the marginal shrinking—through attrition—of the size of the civilian work force (Jacobson, 2018). Not surprisingly, President Trump's approach to performance management, while similar in its practices to Obama's, aspires to a business model, envisioning the citizens as customers and holding federal employees accountable. Margaret Weichert, Deputy Director for Management at OMB, sees a central role for data in this enterprise, with "drivers" of the agenda being information technology, data accountability/transparency, as well as a modern workforce (Clark, 2018, p. 16). But if not for prototypically Trump-like messages signaling disdain for career bureaucrats, advocating streamlined processes to remove poor performers, and pushing back at unions (Katz, 2018, p. 7), this technology-driven emphasis might well have come from Trump's predecessor.

Similarities between Obama and Trump regarding use of data extend to drone strikes as well. The day after inauguration, Trump authorized the use of strikes in Syria though departing from Obama national security processes, which reserved the strike capacity for the Pentagon. Under Trump, the CIA both collects the intelligence that fuel the targeting and carries out the strikes. The turf maneuvering between the Pentagon and CIA could have real ramifications, since they reportedly employ different standards of algorithmic certainty, with the CIA's "near certainty" decision-rule more demanding than the Pentagon's "reasonable certainty" (Lubold

[1] For more information see www.opensecrets.org/pres16/expenditures?id=n00023864

& Harris, 2017). At the same time, empowering the CIA to conduct the strikes removes the process even more so from the scrutiny of congress and the courts, with CIA activities, relative to the Pentagon's, shielded more from view. Put differently, the data and analytics may remain the same, but process differences could have a real impact.

Still, in other areas that move beyond the cases explored in this article, Trump has taken aim at data. He famously banned the Center for Disease Control (CDC) from using terms like "evidence-based" and "science-based" (Sun & Eilperin, 2017), removed data from the website of the Environmental Protection Agency, and disbanded advisory councils that might challenge his own beliefs about the climate and the economy. Furthermore, Trump purged from the web the White House Visitor Log, citing national security risks; the machine-searchable list of visitors to the White House and the Eisenhower Executive Office Building (EEOB) had been made readily available under the Obama administration in the interest of transparency and probably not much of a national security threat. Not surprisingly, given signals on both sides of the question, something of a debate still wages about whether Trump carries on a "war on data".

5. Theoretical Insight Regarding Data and the Presidency

This article's description of data and evidence-based approaches used in the U.S. presidency is necessarily incomplete, dependent on a handful of cases, focusing on one presidency with some insight into another. Indeed, the portrait of the prominence of data in the presidency is if anything modulated by these cases. Performance management is a bit of a sleeper, and the data-driven campaign model, since showcased by Obama, has been diffused widely, across parties, down the ballot and even to campaigns in different international settings, so much so that even it is a little passé.

But Obama did deploy data in far more contexts than described in this article. Technology reporter Nancy Scola dubbed him the "big data president", with some eighty-five big data projects ongoing in his time in the executive (Scola, 2013). Under Obama's watch, the National Institutes of Health (NIH), worked to facilitate delivery of healthcare targeted to a patient's unique genetic makeup, with the goal of collecting data from one million volunteers (NIH, n.d.). The Justice Department (2016), in conjunction with the Whitehouse and law enforcement agencies, released data on police actions, "to increase transparency and accountability and build trust with...communities". And as documented—along with other projects—by the Executive Office of the President (2012), "Mission-oriented Resilient Clouds" would detect and respond to security threats in cloud computing.

The allure of these programs—as well as that of the cases described in this article—is readily apparent, given their stated ambitions. And precisely because of natural allure we should proceed warily, to remember that a critical lens would caution against ignoring the underlying assumptions and power relationships that undergird the processes related to data. With this in mind, this article returns to the three cases, introducing a focus on the political economy of the data.

5.1. A Presidential Data "Revolution"

At first blush, the data-heavy model of campaign politics, performance management and drone strikes under Obama all entail a conceit that the practices are democratizing. After all, a microtargeting process resulting in direct voter contact—campaign personnel reaching out on the phone and at the door—is a decided departure from the mass media model of campaigns that had become prominent over the final decades of the twentieth century. The democratic nature of performance management in the executive is a little different, but it entails the ability of the electorate, as mediated by representatives in Congress, to hold the vast unelected bureaucratic state accountable. Even signature drone strikes, obscured from the view of the public and most elected officials, arguably have an attendant democratic sense. Protecting the U.S. military from ground combat, as uncertain as that is up against non-state opponents in the fight against terrorism, is democratically significant in that military personnel are disproportionately drawn from lower economic classes.

In each of these cases, the data and analytics operations exploited new technologies. Granted microtargeting was around long before Obama (Malchow, 2003), and even the basic architecture of the data employed by Obama was already in place (Hersh, 2015; Kreiss, 2016). But the extent of processing power and the servers, especially in the Catalist world of data mining and modeling with essentially a universe of cases, were fundamentally new and characteristics of big data. In contrast, the technology of data collection in performance management was not cutting edge, but the dissemination of data was. The Obama Administration prioritized the distribution of data, with its performance.gov portal, along with the heralded data.gov portal, which in late October 2015 offered some 189,000 data sets and as of late 2018 over 300,000 data sets to the public. As for signature drone strikes under Obama, it wasn't so much the new tool of drones, because unmanned aerial vehicles have a long history. It was the ability to equip the operation with digital and mobile data, abundant on the ground, then mined, integrated and analyzed to inform algorithmic decisions.

These three data applications, beyond holding democratic allure and the draw of new technology, have the added appeal of secondary instrumental benefits. Regarding drones strikes, a variety of polling data shows that the American public is not particularly critical of them, and presumably successful strikes, not putting American personnel directly at risk, secures stronger public support. A similar byproduct accompanies performance management. Early applications of the practice did result in an upswing of trust of government (Kamarck, 2013). And political campaigns realize multi-faceted instrumental benefits from data and metric-driven efforts. The metrics are used to motivate volunteers and staff and to hold them accountable. Even more so, they serve as concrete and persuasive evidence for donors of the impact and promise of the campaign in the absence of more definitive measures like election outcomes.

It may only be a slight exaggeration to suggest these uses of data in the presidency hold some apparent revolutionary potential. Not only do the data contribute to a desired outcome, but they purport to even change usual power dynamics—offering voice to those not typically heard, as well as a new role for or protection of the average American. But this same revolutionary potential makes it easy to glance away, ignoring inherent biases and threats associated with data in general and more specifically in the presidency.

5.2. The Political Economy of Presidential Data

Three common threads are woven through the cases examined, and they expose concerns that deserve to be addressed in both decisions to turn to data and evaluation of success. The first of these is the uncritical acceptance of efficiency as a goal, most directly borne out in performance management, with its emphasis on outcomes and the use of metrics to judge success.

The approach to performance management manifest in government used by of Obama—Trump too—and predecessors was adopted initially from the private sector in the 1970s. The "New Public Management" aspired to "create market like conditions within the government…to run them 'more like a business'" (Muller, 2018, p. 51). But despite numerous shortfalls of this market application in governing—like metrics distorting incentives and representing overly simplistic conceptions of what motivates personnel—the practice was well-entrenched by the late 1990s (Muller, 2018, p. 55) and it continues today. Muller (2018) finds that metrics that drive management often operate perversely, drawing attention to only things that can be measured and even stifling innovation. There is, however, even a more fundamental concern with this business model, in that it poses efficiency as a preeminent goal.

Efficiency is a common metric employed in the private sector, but its adoption in politics and governing may be at the expense of other things valued. Microtargeting in campaigns, for example, is premised on the efficient deployment of resources to mobilize and persuade enough voters to win an election. But the flip side of targeting voters is that some are ignored, deemed either lost causes or even certain supporters, neither warranting attention by the campaign. And while this may effectively carry a candidate across the line in a given election, it represents a narrow, short-term focus that may not contribute to building an electorate that will support the party in the future. Sociologist Robert Merton called this "the imperious *immediacy* of interests" (Merton, as cited in Muller, 2018, p. 170, emphasis added), wherein individuals look only as far as the short-term consequences of their action.

Zeynap Tufekci's (2012) problem with the efficiency in campaigns is a little different, namely that they *will succeed* in efficiently engineering the electorate. Tufekci is first concerned that the "scalpel" of microtargeting is deployed in private, not subject to public scrutiny. But the bigger problem is that it just may be effective, especially for well-financed campaigns with the resources to devote to data and persuasive techniques. Even if not effective—even if the data which guide the appeals are flawed and replete with errors, as any staffer or volunteer who has worked with these data knows—that campaigns are treating the electorate as a target of their engineering efforts is itself a cause for worry.

Data enterprises that posit efficiency as a goal is a first thread that runs through the cases. A second thread is that the data-based presidency is inexorably tied to a private sector that both supports and benefits from it. The interface and sometimes the data that the campaigns use are held in private hands. NGP-VAN and Catalist, the left's go-to data interface and source of mined data respectively, along with thousands of other paid vendors, constitute the for-hire network of data professionals, many of whom move back and forth between the campaigns and the party apparatus from election to non-election seasons (Kreiss, 2016). It's notable that the combination of the Democratic data and the privately-held NGP-VAN interface, according to Kreiss (2016), serves as a "robust piece of infrastructure that the party's technology ecosystem convenes around". In other words, the private data actually structure the party organization. And in a related fashion, the dependence of a political party and its campaigns on a small number of private firms cannot help but affect where power rests within the party organization, not necessarily with the voter or the party elites, but with vendors.

Performance management, like the data and analytics in campaigns, is subject to a revolving door of sorts regarding leadership. While not universal, a common pattern is that top personnel responsible for performance management, and OMB directors as well, are drawn from the private sector—or at least from those with experience in the private sector. It's also the case that many of these management leaders return then to the private sector after service. Admittedly, the career of Shelley Metzenbaum, President Obama's "performance guru", was more entangled with academia and other governmental positions than the private sector. But Trump performance management leader Margaret Weichert demonstrates a clear trajectory into government from the private sector.

The undercurrent of values and practices that inject a market-based influence into data in the presidency extends to a third dimension as a well: the transaction marking the exchange of data. This is evident in the data used by campaigns and in signature strikes, with the first involving an implicit transaction with the state and the second with individuals, granted, many outside of the boundaries of the U.S.

The campaign data originate in lists of registered voters compiled by the U.S. states. As a condition of voting, individuals provide data to the state, but then private firms like NGP-VAN offer user-friendly tools for working with the data to parties and campaigns. Or in the case of Catalist, the data are augmented through integration with other sources, using algorithmic processes to add synthetic measures (Hersh, 2015). In the abstract, these data and tools, the likes of which are used by presidential candidates, represent an implicit transaction, one in which data collected by the state and made available at little cost is collected by businesses, then sold in a repackaged form to political organizations. Of course, compensating the intermediaries for the value added to the data seems only right. Yet it introduces the question of whether there is just compensation for the original data provider.

For Phil Howard (2018) the answer is "no", at least with reference to the big social media players and political mining firms like Cambridge Analytica, which extract data with ease. Howard is concerned that the citizens have no effective control over their data, which will be used for political purposes. Among the mix of Howard's recommendation to put some degree of control back in the hands of the public is that individuals should be able to donate their data to "the civic groups, political parties, or medical researchers they want to support" allowing them to leverage their own data for political purposes.

This transactional sense of data emerges in Evgeny Morozov's (2017) analysis of artificial intelligence (AI) as well. Morozov asserts that the compensation received by individuals for the data that fuel AI research and development—compensation that is nothing more than access to a social network—is modest when considering the price that government and individuals will pay for products created by AI. Signature strikes invoke an element of this same transactional logic, with a troubling addition. The data that fuel the strikes represent the digital exhaust of users on the ground, snatched up by surveillance operations. In this, access to internet and mobile technology is the compensation for the user, which admittedly may be of substantial value. But that the data are then deployed to target for the purpose of killing individuals with only some stated degree of statistical certainty, and that this practice captures innocent bystanders as well, has an element of perversity to it. Admittedly, national security and covert operations are not the same as AI enterprises, and it's absurd to suggest that those being surveilled should be better compensated for their data. But this transactional calculus regarding signatures strikes, just like those implied or described by Kreiss (2016), Howard (2018) and Morozov (2017), at a minimum, points to the merits of looking well beyond the effectiveness of the data and processes as measured by numbers of terrorist killed.

5.3. Moving Forward

This article has suggested that the world of data—including big data—is borne out in the U.S. presidency, in some cases accompanied by the buzz that this is fundamentally new, even to the point of revolutionary, potentially disruptive of traditional power arrangements. But the subtle irony is that viewed through a critical lens, those traditional power arrangements may prevail, in some cases enhanced by the perceived revolutionary potential of the data and data-related processes used in the presidency.

But one need not focus on the political economy of data to identify ways in which the popular understanding of data and the potential they hold are entangled with political and ethical concerns. Consider the alarm generated by looking closely at algorithms, challenging the conceit that they are immune from prejudice. Cathy O'Neil (2016) demonstrates how algorithmic-informed decisions can reinforce the existing biases of society, that policing tools using predictive modeling carry the appearance of objectivity but can be "tools of math destruction", perpetuating existing traditional class biases. What's more, data journalism outlet ProPublica offers a telling rejoinder to statisticians who judged as fair the algorithms used in sentencing recommendations, finding that when applied to actual people, the algorithms systematically overestimate the threat of recidivism for African American defendants and underestimate it for White defendants (Caplan, Donovan, Hanson, & Matthews, 2018).

Ethical concerns come into the picture typically in ways subtler than the life/death calculations marking signature drone strikes, and because of this they slip by undetected, especially when safety procedures don't intervene successfully. Cambridge Analytica's misuse of data was facilitated by Facebook permitting academic Aleksandr Kogan to harvest its user data, despite Kogan's research proposal being rejected by his university's ethics board (Weaver, 2018). Even with a definitive say by an Institutional Review Board, these bodies tend to gravitate toward a legalistic review of proposals, and—furthermore—are frequently ill-equipped to tease out the ethics of big data (Metcalf, Keller, & boyd, n.d.). It goes without saying that many of the decisions regarding data escape scrutiny by experts tasked with reviewing ethics, especially in the realm of politics and government.

This is all to say that from a number of perspectives—whether viewing data deliberately though a critical lens or simply examining the current areas of concern regarding data and big data—it's clear that the U.S. presidency faces substantial data-related challenges. Despite the occasional utterance of Trump that calls expertise into question, the U.S. political system continues to value information and expertise, both of which contribute to the political capital of actors and institutions. And the norms of science still prevail, indeed even with some new role for social-scientific applications in the political world. In short, there is no reason to believe that "data-driven" is a passing phase. But it's time to contemplate what a measured approach to data would look like. To be concrete, the goal should be to deploy data in effective and ethical ways, all the while alert to the biases and shortcomings that underlay their collection and use. This is no small task, especially in an environment that routinely prioritizes quick action over deliberation, particularly on matters that may require extraordinary technical expertise, though sound and detached political judgment as well.

Acknowledgments

The author thanks colleagues Michael Guenther (History), Janet Seiz (Economics) and Eliza Willis (Political Science) for sharing their insight.

Conflict of Interests

The author declares no conflict of interests.

References

Becker, J., & Shane, S. (2012, May 29). Secret 'kill list' proves a test of Obama's principles and will. *The New York Times*. Retrieved from www.nytimes.com

Caplan, R., Donovan, J., Hanson, L., & Matthews, J. (2018). Algorithmic accountability: A primer. *Data & Society*. Advanced online publication. Retrieved from datasociety.net/output/algorithmic-accountability-a-primer

Clark, C. (2018). Improving performance (in Trump's government makeover). *Government Executive*. Retrieved from www.govexec.com

Engage. (2012). Inside the cave. *Engage*. Retrieved from enga.ge

Executive Office of the President. (2012). *Big data across the federal government*. Washington, DC: Executive Office of the President. Retrieved from obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/big_data_fact_sheet_final_1.pdf

General Accountability Office. (2013). *Managing for Results: 2013 federal managers' survey on organizational performance and management issues* (GAO-13-519SP Report). Washington, DC: General Accountability Office.

Green, J., & Issenberg, S. (2016, October 27). Inside the Trump bunker, with days to go. *Bloomberg*. Retrieved from www.bloomberg.com

Greider, W. (1992). *Who will tell the people?* New York, NY: Simon & Schuster.

Harris, S. (2015). Managing for social change: Improving labor department performance in a partisan era. *West Virginia Law Review*, *17*, 100–158.

Hersh, E. (2015). *Hacking the electorate: How campaigns perceive voters*. New York, NY: Cambridge University Press.

Howard, P. (2018, July 16). Our data, ourselves. *Foreign Policy*. Retrieved from foreignpolicy.com/2018/07/16/our-data-ourselves-democracy-technology-algorithms

Hu, M. (2017). Metadeath: How does metadata surveillance inform lethal consequences? In R. Miller (Ed.), *Privacy and power: A transatlantic dialogue in the shadow of the NSA-affair*. Cambridge: Cambridge University Press.

Issenberg, S. (2012a). *The victory lab*. New York, NY: Crown Publishers.

Issenberg, S. (2012b, May 22). The death of the hunch. *Slate*. Retrieved from www.slate.com

Jacobson, L. (2018, January 22). Taking the measure of the federal workforce under Donald Trump. *Politifact*. Retrieved from www.politifact.com/truth-o-meter/article/2018/jan/22/taking-measure-federal-workforce

Jochum, E. (2009, September 24). OMB will create new performance management framework for agencies. *Government Executive*. Retrieved from www.govexec.com

Justice Department (2016). Growing number of communities are using data to improve policing and criminal justice. *Department of Justice Archives*. Retrieved from www.justice.gov/archives/opa/blog/growing-number-communities-are-using-data-improve-policing-and-criminal-justice

Kamarck, E. (2013). Testimony before House Committee on oversight and government reform: Lessons for the future of government reform. *Governance Studies at Brookings*. Retrieved from www.brookings.edu

Katz, E. (2018). Modernizing the workforce (in Trump's government makeover). *Government Executive*. Retrieved from www.govexec.com

Kreiss, D. (2016). *Prototype politics: Technology-intensive campaigning and the data of democracy*. New York, NY: Oxford University Press.

Kruzel, J. (2018, March 16). Why Trump appointments have lagged behind other presidents. *Politifact*. Retrieved from politifact.com

Lubold, G., & Harris, S. (2017, March 13). Trump broadens CIA powers, allows deadly drone strikes. *Wall Street Journal*. Retrieved from www.wsj.com

Malchow, H. (2003). *The new political targeting*. Washington, DC: Campaigns and Elections.

Mayer-Schöenberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we life, work and think*. Boston, MA: Houghton Mifflin Harcourt.

Metcalf, J., Keller, E., & boyd, d. (n.d.). Perspectives on big data, ethics, and society. *Council for Big Data, Ethics, and Society*. Retrieved from bdes.datasociety.net

Morozov, E. (2017). *Do we have a right to our data? Data ownership and the inequality debate*. Public address

at Grinnell College, USA.
Muller, J. (2018). *The tyranny of metrics*. Princeton, NJ: Princeton University Press.
National Institutes of Health. (n. d.). About the All of Us research program. *National Institutes of Health*. Retrieved from allofus.nih.gov/about/about-all-us-research-program
O'Neil, C. (2016). *Weapons of math destruction*. New York, NY: Crown.
Orren, K., & Skowronek, S. (2017). *The policy state: An American predicament*. Cambridge, MA: Harvard University Press.
Rosenberg, M., Confessore, N., & Cadwalladr, C. (2018, March 17). How Trump consultants exploited the Facebook data of millions. *The New York Times*. Retrieved from www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html
Scahill, J., & Greenwald, G. (2014, February 10). The NSA's secret role in the U.S. assassination program. *The Intercept*. Retrieved from theintercept.com/2014/02/10/the-nsas-secret-role
Scola, N. (2013, June 14). *Sizing up the executive branch: Fiscal year 2017*. Washington, DC: United States Office of Personnel Management. Retrieved from www.opm.gov/policy-data-oversight/data-analysis-documentation/federal-employment-reports/reports-publications/sizing-up-the-executive-branch-2016.pdf
Sun, L., & Eilperin, J. (2017, December 15). CDC gets list of forbidden words: Fetus, transgender, diversity. *The Washington Post*. Retrieved from www.washingtonpost.com/national/health-science/cdc-gets-list-of-forbidden-words-fetus-transgender-diversity/2017/12/15/f503837a-e1cf-11e7-89e8-edec16379010_story.html?noredirect=on&utm_term=.5bc086fc9789
Susser, B. (1992). *Approaches to the study of politics*. New York, NY: Macmillan.
Tufekci, Z. (2012, November 16). Beware the smart campaign. *The New York Times*. Retrieved from www.nytimes.com/2012/11/17/opinion/beware-the-big-data-campaign.html
USDA. (2013). *FY 2013 annual performance report*. Washington, DC: USDA. Retrieved from www.ocfo.usda.gov/docs/FY%202013%20Annual%20Performance%20Report.pdf
Vogel, K., & Samuelsohn, D. (2016a, January 5). Trump quietly builds a data juggernaut. *Politico*. Retrieved from www.politico.com
Vogel, K., & Samuelsohn, D. (2016b, June 28). Trump's secret data reversal. *The New York Times*. Retrieved from www.nytimes.com/2012/11/17/opinion/beware-the-big-data-campaign.html
Weaver, M. (2018). Cambridge University rejected Facebook study over 'deceptive' privacy standards. *The Guardian*. Retrieved from www.theguardian.com
Zenko, M. (2012a). *Daniel Klaidman's revelations*. New York, NY: Council on Foreign Relations.
Zenko, M. (2012b). *Targeted killings and signature strikes*. New York, NY: Council on Foreign Relations.
Zenko, M. (2013). *Reforming U.S. drone strike policies* (Special Report 65). New York, NY: Council on Foreign Relations.

About the Author

Barbara Trish is Professor of Political Science at Grinnell College (Grinnell, IA), where she also directs the Rosenfield Program in Public Affairs, International Relations, and Human Rights. Her analyses of U.S. politics have been published in scholarly journals, edited volumes and in the popular press.

Politics and Governance (ISSN: 2183–2463)
2018, Volume 6, Issue 4, Pages 39–47
DOI: 10.17645/pag.v6i4.1757

Commentary

Cloud Computing in Singapore: Key Drivers and Recommendations for a Smart Nation

Reuben Ng [1,2]

[1] Lee Kuan Yew School of Public Policy, National University of Singapore, 259772 Singapore, Singapore;
E-Mail: spprng@nus.edu.sg
[2] Geriatric Education and Research Institute, 769027 Singapore, Singapore

Submitted: 1 October 2018 | Published: 21 November 2018

Abstract
Cloud computing adoption enables big data applications in governance and policy. Singapore's adoption of cloud computing is propelled by five key drivers: (1) public demand for and satisfaction with e-government services; (2) focus on whole-of-government policies and practices; (3) restructuring of technology agencies to integrate strategy and implementation; (4) building the Smart Nation Platform; (5) purpose-driven cloud applications especially in healthcare. This commentary also provides recommendations to propel big data applications in public policy and management: (a) technologically, embrace cloud analytics, and explore "fog computing"—an emerging technology that enables on-site data sense-making before transmission to the cloud; (b) promote regulatory sandboxes to experiment with policies that proactively manage novel technologies and business models that may radically change society; (c) on the collaboration front, establish unconventional partnerships to co-innovate on challenges like the skills-gap—an example is the unprecedented partnership led by the Lee Kuan Yew School of Public Policy with the government, private sector and unions.

Keywords
Keywords: big data; cloud computing; public management; psychomics; public policy; Singapore; smart city

Issue
This commentary is part of the issue "Big Data Applications in Governance and Policy", edited by Sarah Giest (Leiden University, The Netherlands) and Reuben Ng (National University of Singapore, Singapore).

1. Introduction

The world generates more data every two days than from the dawn of early civilization through the year 2003 combined. And data rates are still growing at approximately 40% per year. Against this background, data has outstripped common warehousing and analytics tools. To move forward, organisations—both public and private—need new capabilities, especially cloud computing, to differentiate their products and services.

Every major study hails the potential of cloud computing and analytics. Gartner found that enterprise spending on cloud computing grew faster than overall IT spending and predicted the technology will grow by over 100% (PRWeb, 2012). Synergy Research Group concluded that in 2016, cloud computing dominated many components within the Information and Communication Technologies (ICT) market, pushing cloud revenue growth above 25% year-on-year. A cloud-enabled business model survey found that 62% of Chief Information Officers and Chief Data Officers consider cloud computing as the top priority for ICT (Berman, Kesterson-Townes, Marshall, & Srivatbsa, 2011). Bessemer Venture Partners reported that the cloud computing market revenue grew 35.8% annually from 2008–2014, and 22.8% annually from 2014 to 2018 culminating in projected revenues of US$127 billion (Figure 1). A 2010 survey found that 23% of Singapore companies adopted cloud technology. More recently, International Data Corporation (IDC) forecasted that the cloud computing market in Singapore will grow to US$1 billion by the end of 2018.

There are three features of cloud computing: first, cloud computing provides services on demand, and these resources are scalable over multiple data centres.

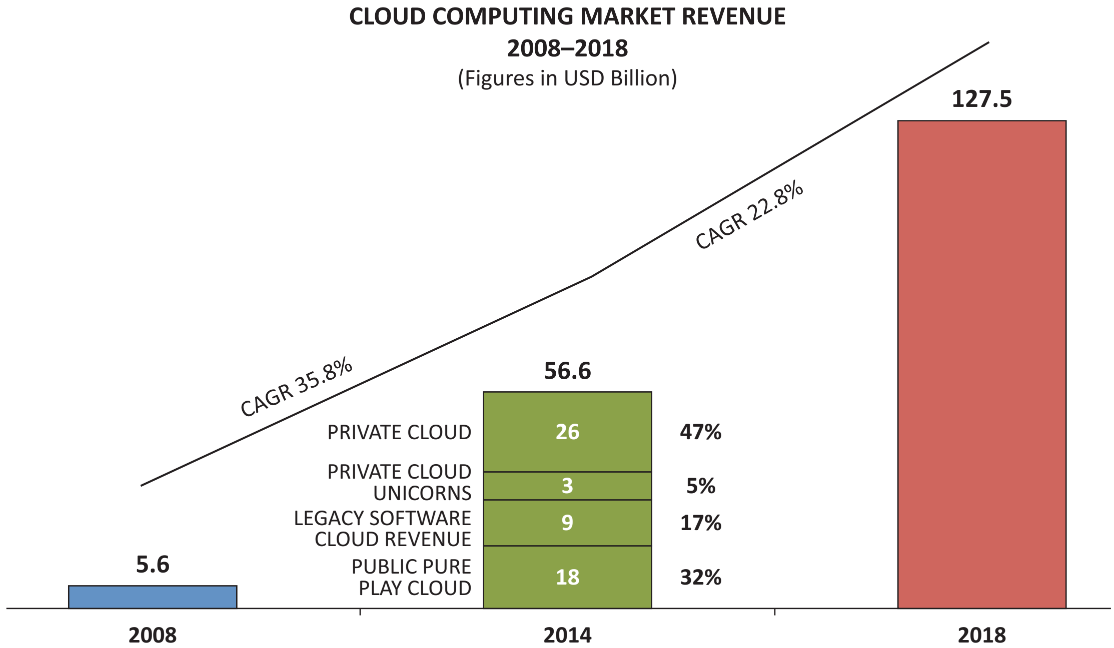


Figure 1. Incredible growth of cloud computing market revenue over a decade (2008–2018; Deeter & Shen, 2015).

Second, these services are easily accessible and location independent, meaning that a user can access the suite of services worldwide through internet browsers. Third, service quality is guaranteed for bandwidth and memory capacity.

These features give rise to three benefits: first, cloud computing provides a more palatable cost structure. It requires no upfront hardware cost compared with traditional data warehouses with high capital investments. The use of cloud computing is recorded as operating expenses and significantly reduced. Second, business risk is reduced as it is outsourced to providers who are better equipped to manage these risks. Third, organisations who are starting their digital journey can leverage on the latest computing technology to leap frog their capabilities compared to a traditional step-wise build up.

Of broader significance, cloud computing is positioned as a crucial enabler in Singapore's Smart Nation plan. The latter is part of a national effort to put Singapore at the forefront of technological innovation and implementation. The Prime Minister took a personal interest in launching the vision in 2014 and in 2017, devoted a significant part of his National Day Rally Speech, an important annual address, to update on the nation's initiatives and aspirations.

The case study will first outline the key drivers for Singapore's adoption of cloud computing. Second, analyse the cybersecurity considerations in cloud computing adoptions. Third, provide recommendations for countries embarking on a journey to enable big data applications in public management and policy.

2. Five Key Drivers of Singapore's Cloud Computing Adoption

2.1. Public Demand for and Satisfaction with E-Government Services

Singapore has done well in ICT infrastructure development as evidenced by the extraordinary high mobile subscription penetration rate of 155.6% and, high internet penetration rate of 135.1% in 2013. In addition, 73% of the Singapore population are internet users, exposed to the information deluge on the world wide web. The statistic is comparable to the average internet users of above 70% in developed countries. On the contrary, fixed telephone line usage has plunged to 36.4% and 25.7% for wired broadband subscriptions. These numbers underscore the high levels of connectivity and mobility of residents in Singapore.

On the public front, through e-government initiatives, Singapore migrated many processes that required face-to-face interactions to self-help channels, along with noteworthy adoption of paperless transactions. As a result, Singapore has been ranked highly in e-government implementation (Table 1), and in 2013, emerged tops ahead of Finland and the US. These initiatives have resonated with the citizenry, garnering positive feedback from 2011 to 2015 from both businesses and individuals, according to 2016 surveys from the Singapore Government Technology Agency (GovTech; Figure 2).

On the government-to-business front, over 1500 representatives from multiple industry sectors surveyed reported that 99% of businesses visited government websites and over 90% were satisfied with the information quality provided, and the ease of completing transactions online.

Table 1. Singapore tops the ranks for e-government implementation in 2013 (Hashemi, Monfaredi, & Masdari, 2013).

Rank	Country	Score
1	Singapore	94.0
2	Finland	93.2
3	USA	93.1
4	Korea	92.3
5	UK	88.8
6	Japan	88.3
7	Sweden	87.8
8	Denmark	83.5
9	Taiwan	83.5
10	Netherlands	82.5
11	Australia	82.1
12	Canada	81.8
13	Switzerland	81.3
14	Germany	80.1
15	Italy	79.1

Similarly, for individuals, 80% visited government websites in the past 12 months and over 90% espoused satisfaction with the ease of locating and comprehending the information provided (Figure 3).

Given the appetite for e-services/platforms and the extraordinary ability of the government to provide them, it is hardly surprising that Singapore ranks highly on cloud readiness—a collection of variables that measure the propensity for cloud adoption. Singapore is ranked ahead of her ASEAN peers in 2011 and inched ahead in 2016 (Figure 4).

The Asian Cloud Computing Association (ACCA)[1] ranked countries in Asia Pacific and Oceania on 10 indicators, which Singapore emerged second behind Hong Kong and ahead of New Zealand (#3), Australia (#4) and Japan (#5) among 14 countries in 2016. Specifically, Singapore scored well in broadband quality, data privacy, government regulation, intellectual property protection, but less well in cybersecurity and freedom of information.

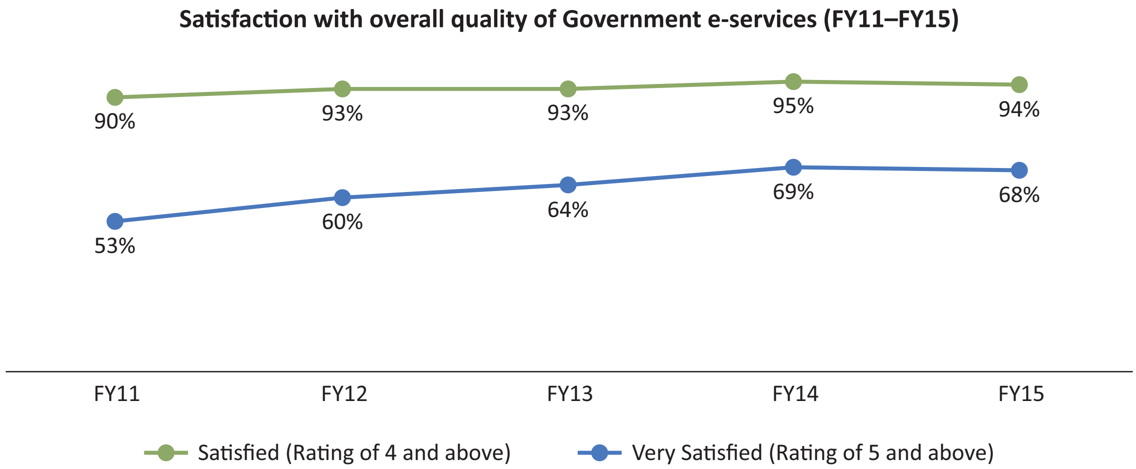


Figure 2. Business users' satisfaction with overall quality of government e-services from 2011–2015.

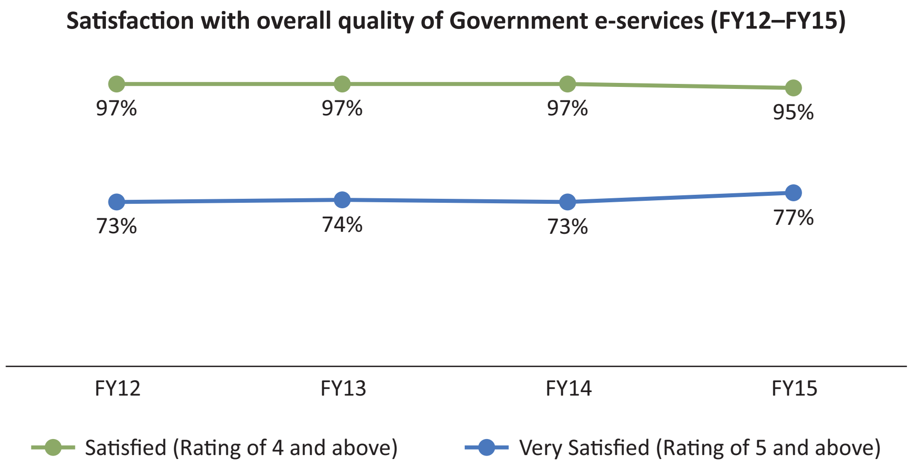


Figure 3. Citizens' satisfaction with overall quality of government e-services from 2012–2015.

[1] More information about the methodology and data sources for the index are available at: www.asiacloudcomputing.org/17-news/306-2016-cloud-readiness-index

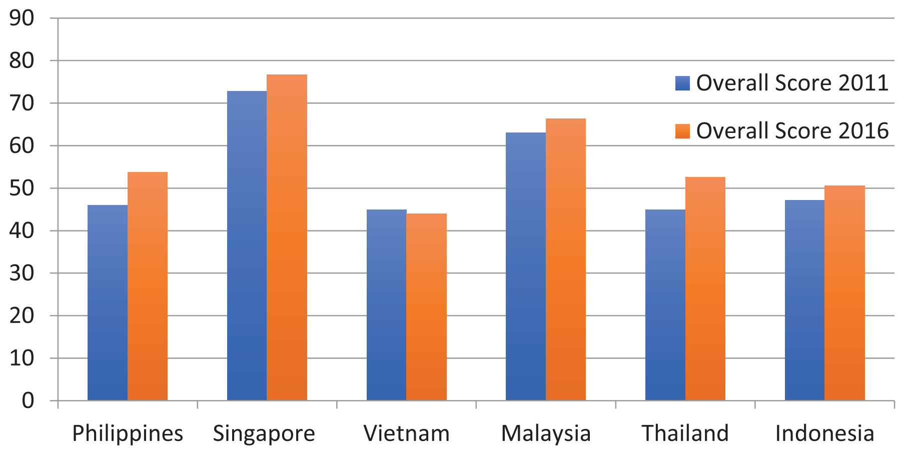


Figure 4. Overall cloud readiness index in ASEAN, 2011, 2016. Source of raw data: Cloud readiness index, 2016.

2.2. Focus on Whole-of-Government Policies and Practices

The public service is moving towards a whole-of-government approach to policy making and operations, providing a key driver for whole-of-government cloud adoption. This approach reflects the reality that many contemporary issues cannot be siloed. For example, ageing is not only a health issue, it also impacts the agencies in transport, environment, social and family. To create opportunities for silos to "talk" to each other, the government organised forums for top leaders within different sectors to discuss and craft policies. For example, the Social Forum brings together senior leaders of all the social agencies to build consensus for social policies. This is also great platform to commission important analytics projects that require data from multiple agencies—a typically arduous process. When these projects are debated and benefits delineated among senior leaders across agencies, they are more willing to share data that contribute to a common good. Such initiatives provide use cases for whole-of-government cloud adoption.

Besides platforms to talk between silos, the Prime Minister's Office started an important team of "bridge builders"—The Strategy Group—to shepherd and coax whole-of-government practices. They staff the sectoral forums, and pioneered the data science commissioning platform to nudge line agencies towards a more whole-of-government approach to policy, practice, and data sharing. At the tactical level, Singapore formed a Municipal Services Office whose app is likened to a one-stop shop for public feedback and manages it across government to ensure a coordinated response. Such centralisations at the strategic and tactical levels are key drivers for cloud computing adoption in the public service.

2.3. Restructuring Technology Agencies to Integrate Strategy and Implementation

To enable whole-of-government innovation and implementation, technology-related agencies are also being restructured. Telco and media regulation are now brought under one agency—the Info-communications Media Development Authority of Singapore (IMDA). Merging both the InfoComm Development Authority of Singapore (IDA), and the Media Development Authority of Singapore (MDA), IMDA will focus on leveraging new technologies for better regulation and application to improves lives.

Since May 2017, the Prime Minister's Office (PMO) started the Smart Nation and Digital Government Office and assigned a Permanent Secretary to lead this group that integrated strategy and implementation capabilities. The strategic capabilities are integrated from the teams formerly in the PMO, Ministry of Finance, and Ministry of Communication and Information. These teams will be responsible for policy and strategy which will be implemented by Government Technology Agency (GovTech) consisting of deep capabilities in data science and software development.

Other developments include bringing together cybersecurity capabilities through the formation of the Cyber Security Agency (CSA) in 2015 to implement the National Cybersecurity Masterplan 2018. On the local front, the Municipal Services Office (MSO) was established in 2015 to coordinate whole-of-government responses to municipal services that were previously managed by different agencies. These major restructuring exercises are both expensive and expansive but are effective to integrate strategy and implementation.

2.4. Building the Smart Nation Platform

A key catalyst of cloud computing is the Smart Nation Platform (SNP) involving a nationwide communications and sensor infrastructure. This platform will enable centralised data aggregation and sharing, as well as provide new capabilities to derive refined insights from cross-referencing a vast array of Government datasets. Datasets can also be shared with the community and industry for co-creation and self-enablement. The SNP will be a key enabler for Singapore to remain on the cutting edge of Government operations and service delivery.

The SNP will connect sensors through the deployment of Aggregation Gateway Boxes (AG Boxes) throughout Singapore. Approximately 2000 AG Boxes could be deployed at major roadside locations and 10,000 AG Boxes at residential estates. AG Boxes provide connectivity to street lights and wireless mesh networks to ensure that wireless sensors can be easily plugged into the network. As AG Boxes are not licensable telecommunication services, the government will own and operate then. Private operators could share the remaining space in each AG box.

The network will carry sensitive data from sensors so the government's ownership of relevant portions of the SNP will assure that the critical components within the network are treated with the highest security and trust. Specifically, the Info-communication and Media Development Authority will own the infrastructure. In doing so, they are not welded to any contractor and can leverage on different vendors to meet changing demands. Private companies will operate licensed services such as the transmission of data and a majority of maintenance.

Essentially, the SNP is projected to carry a significant amount of Government sensor data. Hence, an integrated platform for the collation, sharing, and analytics will ensure more coherent insights and swifter deployment of government services. The cloud computing backbone is the Smart Nation Operating System (SN–OS) that pull together all sensor and other data types for sense-making. The SN-OS consists of three platforms: sensor management, data exchange and sense-making.

Through the SN–OS, public sector entities will be able to access cross-agency sensor data to analysis and decrease duplicity of data collection. An analytics layer will facilitate the merging of different datasets (e.g., sensor and admin) to inform policy positions and research. Underlying all this cloud infrastructure is a data governance framework with proper access restrictions and audit trails to ensure that that only the right officers can access the data.

In sum, the Smart Nation Platform is an ambitious example of a comprehensive and massive cloud computing platform that sets the foundation for whole-of-government policy-making and practices.

2.5. Purpose-Driven Cloud Applications in Healthcare

2.5.1. Healthcare Transformation

Within the healthcare space, cloud computing is applied at scale to healthcare transformation. Signalling the government's commitment to transformation, the Ministry of Health (MOH) recently appointed outgoing National University of Singapore President and former MOH Director of Medical Services, Professor Tan Chorh Chuan, to direct the inaugural Healthcare Transformation Office as Singapore's first Chief Health Scientist.

Setting up this important national function will accelerate the use of cloud computing especially in population health analytics and tele-medicine. Nationally, the healthcare sector was restructured into three clusters where each cluster consists of entities in the continuum of care (e.g., acute, community, primary care and preventative health). Each cluster is responsible not only for acute care through large tertiary hospitals but also population health within the geographic cluster. This involves disease prevention, acute care, chronic disease management, and long-term care.

Each cluster utilises their own cloud which is linked to the common registry, a large private cloud known as the National Electronic Health Records (NEHR). This national database contains demographics, subvention and financial data that can be used to stratify the population according to risk for various health conditions and frequent admission (Ng, Hiew, Goh, & Tan, 2018). New data sources from other government agencies such as social and family relationship, birth and death data are explored to further enrich the registry. Further, the registry is enriched by longitudinal and multi-disciplinary data around each individual's health.

Such a cloud platform yields invaluable insights on the ageing population that will support primary prevention efforts to better manage the size of population with chronic diseases. At the national level, disability projections (e.g., Ng, Lim, Saw, Francis-Tan, & Tan, 2018) are done, along with predictive modelling to stratify the population and predict the propensity of groups to develop pre-identified medical conditions such as heart diseases and diabetes.

2.5.2. Tele-Medicine

Another application of cloud computing is tele-medicine. The aim of tele-health is to shift from institution-based care towards home and community care, augmenting healthcare resources (e.g., allied health professionals, mental health psychiatrists etc.) in the system. Essentially such technologies shift from a doctor-centric model to a team-based model to manage the shortage of doctors and promote holistic care.

Tele-health applications bring cost-effective supervised rehabilitation to older homebound patients who face significant challenges of traveling to outpatient rehabilitation centers. By exploiting the latest technologies to develop rehabilitation devices specially designed for home and remote use, homebound patients do not need to travel to outpatient rehabilitation centres. Such video-conferencing technology will also help patients to seek advice and guidance from all members of the multidisciplinary rehabilitation team. Tele-rehabilitation will be cost-effective as the rehabilitation team will not need to travel to homebound elderly homes.

Of broader significance, tele-medicine creates a platform for the transmission of real-time data. This has two benefits. One, the rehab team could monitor patients more intensively and calibrate the rehab exercise more appropriately. Two, with cloud computing enabled tele-monitoring, care-givers can remotely track elderly citizens or immobile individuals living alone through home/wearable sensors to detect and respond early to unusual extended periods of non-movement or potential falls. Consumers can utilise Telehealth services from their home/ community for greater access to health education and to obtain timely proper care, which has been previously confined to specialised healthcare settings.

The former Info-communication Development Authority of Singapore (IDA) is rolling out the infrastructure for nationwide ultra-high-speed broadband access of 1Gbps and more, known as the Next Generation Nationwide Broadband Network, to all physical addresses including homes, schools, government buildings, businesses and hospitals. Such infrastructure will support the implementation of a nation-wide tele-rehabilitation program for patients.

Another cloud computing application is a commercial app to link patients and doctors by a Singapore-based company, RingMD that manages 1.5 million patients in over 50 countries. The firm aims to provide healthcare to under-served population by accessing the best doctors through tele-consultations. This is predicated on high mobile phone penetration rates in both developed and developing countries, enabling these consultations through mobile phones and devices.

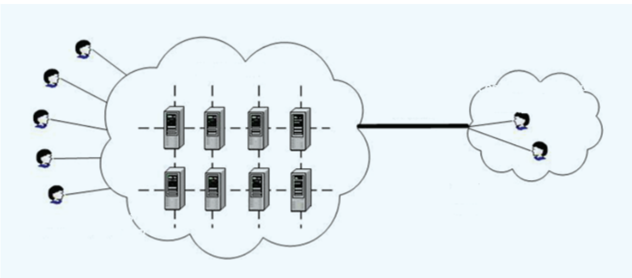
Users enrol on the RingMD platform and consultations are done through a video link on mobile devices. Individuals can also wear devices that transmits their vital signs to doctors' in real time. Conditions that do not require physical checks can be remotely diagnosed and treated. Through analysing copious amounts of data, global and local insights can be provided to patients, doctors and caregivers.

In sum, the key driver for cloud computing in Singapore is purpose-driven, especially in healthcare through healthcare transformation and tele-medicine. With the formation of a new healthcare transformation office driving population health analytics, and a bustling commercial health sector, the adoption of cloud computing is projected to accelerate.

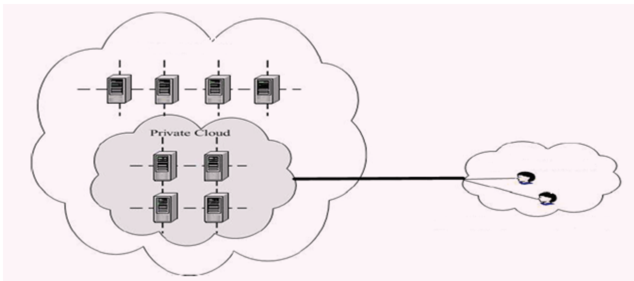
3. Cybersecurity and Cultural Considerations in Singapore's Cloud Adoption

There are four models of cloud adoption (Figure 5): public cloud, as the name suggests, is accessed by the general public. An example is the government sharing portal data.gov.sg where government datasets are shared for co-creation and public consumption. The use of public cloud benefits from the lower computing cost. Community Cloud serves a sector and is best exemplified by the Singapore Ministry of Education's iCONnect system, an e-mail and collaboration platform for the teaching profession. Both the public and community clouds are within the medium assurance zones where computing resources are shared with different cloud users at decreased cost.

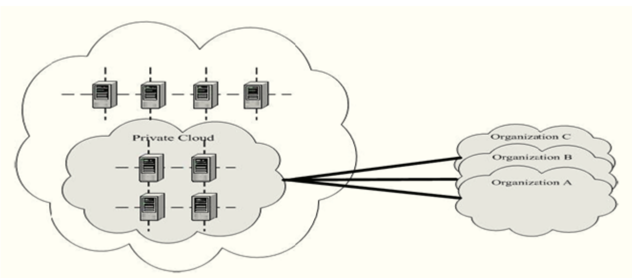
A private cloud is used for most of government agencies. Known as the Central G-Cloud, it meets high assurance needs of providing a dedicated computing resource



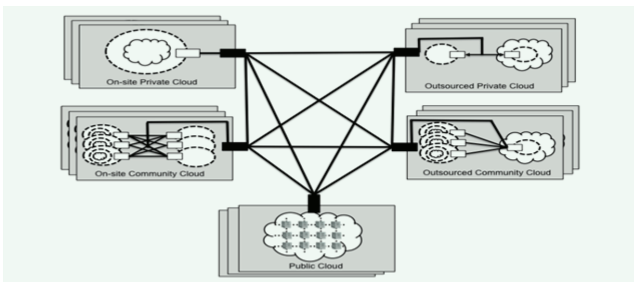
Public cloud. The cloud infrastructure is assessed by the general public or a large industry group.



Private cloud. The cloud infrastructure is operated solely for an organization.



Community cloud. The cloud infrastructure is shared by several firms within a sector with shared considerations (e.g., security, compliance, etc).



Hybrid cloud. The cloud infrastructure is a composition of two or more clouds (private, community or public). Though these respective clouds have unique properties, they have common applications.

Figure 5. Four models of cloud adoption (Ernst & Young Global Limited, 2014).

within government. Most web service exchange and gateways leverage the G-Cloud. The predominant use of a private cloud is driven, among other reasons, by cyber-security and a prevention-focused cultural tendency.

Cross-cultural experts who categorise national cultures have often described Singapore as "tight", characterised by strong social norms and low patience for deviant behaviour. In a study of 33 countries, Singapore emerged in the top quartile of tightness scores. In other studies, Singapore was categorised as a prevention-focused society where civil servants are detail-orientated and place great emphasis on the absence of *negative* outcomes, perpetuating a state of vigilance where individuals are careful not to make mistakes (Brown, Abdallah, & Ng, 2011; Higgins et al., 2001), and high on pettiness (Ng & Levy, 2018). The opposite is that of a promotion-focused society where the focus is on nurturance, and the presence of *positive* outcomes (Hamamura, Meijer, Heine, Kamaya, & Hori, 2009). These empirical studies on culture provide an indication of the mindset of the government in employing a private cloud.

More importantly, a private cloud attempts to mitigate cyber-security risks. As seen in Figure 6, Singapore's cyber-security prowess is not as strong compared with Australia, South Korea and India according to the Cloud Readiness Index 2016 compiled by the ACCA. In new world of data explosions, information security is of paramount consideration. There is intense competition between cryptographers attempting to protect data and hackers attempting to steal it. When the latter happens, public trust is undermined. With Singapore's push to build the world's first Smart Nation, an unprecedented number of devices are being connected to the internet, escalating the risk of expansive disruptions in cyberspace, potentially disrupting defence, public transport and stock markets.

Exacerbated by the Internet of Things, the Cambridge Analytica scandal (Ng, 2018), the accessibility and decreasing cost of hacking tools, have brought about increased risks, enabling lone-wolfs and terrorists to wreck cyber havoc. Over 5% of American organisations lost north of US$1 million in 2013 to cyber-crime. More staggering, managing the fallout from hacks costs the world about US$445 billion per year. Further, the obscure origins of cybercrime make it difficult to differentiate between state and non-state actors. The challenges that cyber-enabled threats bring for Singapore will be non-trivial.

Against the background, Singapore embarked on a coordinated approach to deal with cyber-enabled threats at the government level. Cyber security is transiting from prevention to response and recovery, with the not-if-but-when mindset. Infrastructure development and policy planning also consider resilience and incident-response capabilities. This is exceptionally paramount in critical information infrastructure (CII) sectors such as banking, transport, health, energy, and telecommunications.

Taken together, the increasing barrage and sophistication of cyber-attacks, a prevention-focused mindset, and the imperative to protect against more damaging cyber-attacks have pushed the government to adopt a private G-Cloud. Though the costs are significantly higher, the Singapore government has probably calculated that the benefits outweigh the cost; more specifically—the potential cost of not implementing a private G-Cloud outweigh the actual cost of doing so.

4. Recommendations for Cloud Computing Adoption in Big Data Applications

Learning from Singapore's advances in the cloud computing journey, we propose the following recommenda-

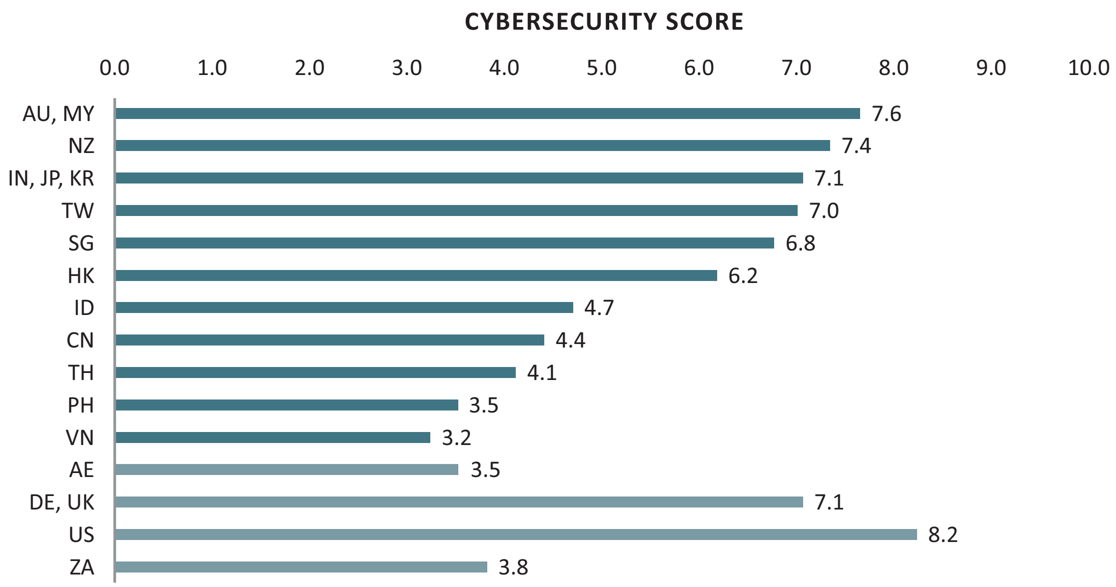


Figure 6. Cybersecurity: Singapore vs. ASEAN (2016). NZ (New Zealand), HK (Hong Kong), TW (Taiwan), JP (Japan), AU (Australia), SG (Singapore), KR (South Korea), PH (Philippines), TH (Thailand), ID (Indonesia), MY (Malaysia), VN (Vietnam).

tions for cloud computing adoption: Technology, regulation and partnerships. Technologically, embrace cloud analytics, and explore "fog computing"—an emerging technology that enables on-site data sense-making before transmission to the cloud. Promote regulatory sandboxes to experiment with policies that proactively manage novel technologies and business models that may radically change society. Seek out unconventional partnerships, beyond the PPP model, to co-innovate on the initiatives to address perennial issues like the skills-gap.

4.1. Accelerate Cloud Analytics and Explore Fog Computing

On the technological front, the current phase for cloud computing is management and organisation with real-time data that are high in volume, velocity and variety. The next phase will be the ability to make sense of the endless streams of data and provide on-demand analytics (Figure 7). While there are fore-runners in the cloud management space, few have been distinguished as market leaders in cloud analytics. Countries can attract up-and-coming firms in cloud analytics and invite them to advance their research and operations.

Fog Computing is another technological advancement to explore. Singapore's Smart Nation and other smart city initiatives around the world have led to many Internet-of-Things (IOT) nodes that ingest data but unable to perform analytics. Cloud servers are too far away to analyse data and respond real-time when required. By 2020, IDC projects that 10% of the world's data will be produced by these devices at the edge. Fog computing is an edge layer with analytics and artificial intelligence (AI) capabilities. The advantages are analytics-on-site and the transmission of only relevant data to cloud servers, driving down costs.

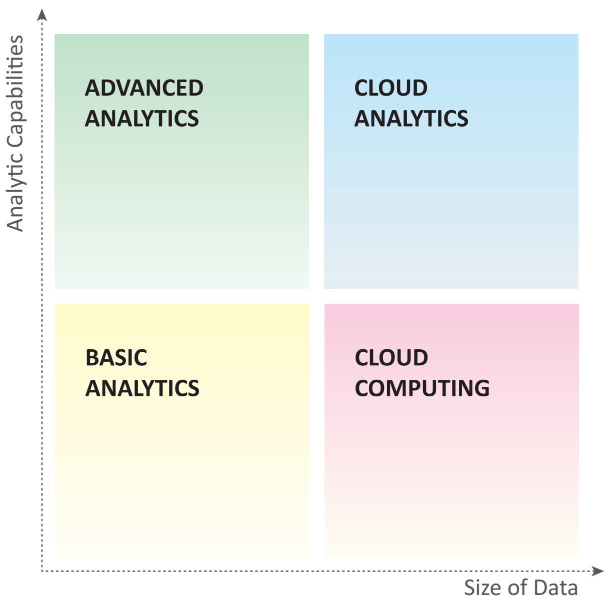


Figure 7. Cloud analytics is a key capability given that data volume and the need for sense-making is increasing exponentially (Hamilton, 2012).

4.2. Create Regulatory Sandboxes for Policy Experimentation

On the regulatory front, the speed of technological advancements has resulted in changes in business models and societal behaviour that are too rapid for policy/regulations to manage. Often, policies and regulations react rather than proactively manage emerging changes, and may hinder the flourishing of new technology. The Monetary Authority of Singapore (MAS), the nation's central bank innovatively created regulatory sandboxes—spaces for policy experimentation—as a means to co-innovate on policies and practices for the explosive FinTech sector. In the sandbox, regulations are relaxed to allow the experimentation of promising fintech products (Ng, 2017). If the outcomes are successful, both the service and policy innovations can be scaled up and implemented. Such incubation practices can be applied beyond the financial sector for policy innovations in the health, social and transport sectors. Further, public perceptions and stereotypes can be measured at different milestones using computational linguistics (e.g. on age stereotypes; Ng, Allore, Trentalange, & Levy, 2015) to shape policy communications.

4.3. Establish Unconventional Partnerships to Co-Innovate on Challenges Like the Skills-Gap

With the gravitation of tech jobs towards certain talent ultra-hubs, there is an urgent need to develop a base with deep and transferable skills for a city aspiring to be such a hub. The prediction of future skills is challenging; the training of a critical mass on these skills to meet future demand is even more challenging. Institutions need a lead time to design new training course that respond to market demands. When these courses are designed and students enrolled, the demand may have shifted.

Attempting to tackle the skills-gap in Singapore, the Lee Kuan Yew School of Public Policy at the National University of Singapore, signed a Memorandum of Understanding (MOU) with government agencies (SkillsFuture Singapore at the Ministry of Education), Technology firms (Microsoft and Linkedin), and the union (National Trade Union Congress). This MOU seeks to mine public and public-sector data to understand the supply and demand of skills within different occupations. Such insights empower individuals to skill up appropriately, training institutions on the types of courses to offer to meet future demand, and government on programs to meet sectoral demand.

Importantly, this MOU represents an unprecedented multilateral partnership between different parties with unique resources to contribute towards addressing a sticky problem such as the skills-gap. Such a partnership across ASEAN could be leveraged to engender a collaborative approach to address a host of labour and social issues that accompany the fourth industrial revolution.

5. Conclusion

This is one of the first known analysis of cloud computing in Singapore with the emphasis of adoption for big data applications in governance and policy. We identified five key drivers: (1) public demand for and satisfaction with e-government services; (2) focus on whole-of-government policies and practices; (3) restructuring of technology agencies to integrate strategy and implementation; (4) building the Smart Nation Platform; (5) purpose driven cloud applications especially in healthcare. These key drivers could serve as learning points and considerations for other nations embarking on their cloud journey.

Conflict of Interests

The author declares no conflict of interests.

References

Berman, S., Kesterson-Townes, K., Marshall, A., & Srivatbsa, R. (2011). *The power of cloud: Driving business model innovation*. Somers, NY: IBM Institute for Business Value. Retrieved from http://www.ibm.com/cloud-computing/us/en/assets/power-of-cloud-for-bus-model-innovation.pdf

Deeter, B., & Shen, K. (2015, June 18). The state of the cloud—2015. *Bessemer Venture Partners*. Retrieved from https://www.bvp.com/blog/state-cloud-2015

Hamilton, B. A. (2012). *Cloud analytics playbook*. McLean, VA: Booz Allen Hamilton. Retrieved from https://www.slideshare.net/BoozAllen/1111212-cloud-playbook-digital

Brown, J., Abdallah, S. S., & Ng, R. (2011). Decision making styles East and West: Is it time to move beyond cross-cultural research? *International Journal of Sociology and Anthropology*, *3*, 452–459.

Ernst & Young Global Limited. (2014). *Building trust in the cloud*. London: Ernst & Young Global Limited. Retrieved from http://www.ey.com/Publication/vwLUAssets/EY_-_Building_trust_in_the_cloud/$FILE/EY-grc-building-trust-in-the-cloud.pdf

Hamamura, T., Meijer, Z., Heine, S., Kamaya, K., & Hori, I. (2009). Approach–avoidance motivation and information processing: A cross cultural analysis. *Personality and Social Psychology Bulletin*, *35*, 454–462.

Hashemi, S., Monfaredi, K., & Masdari, M. (2013). Using cloud computing for e-government: Challenges and benefits. *International Journal of Computer, Information, Systems and Control Engineering*, *7*(9), 596–603.

Higgins, E. T., Friedman, R. S., Harlow, R. E., Idson, L. C., Ayduk, O. N., & Taylor, A. (2001). Achievement orientations from subjective histories of success: Promotion pride versus prevention pride. *European Journal of Social Psychology*, *31*, 3–23.

Ng, R. (2017, March 29). Learning the art of implementing analytics. *Business Times*. Retrieved from https://www.businesstimes.com.sg/opinion/learning-the-art-of-implementing-analytics

Ng, R. (2018, June 28). Learning from the Cambridge Analytica fiasco. *Business Times*. Retrieved from https://www.businesstimes.com.sg/opinion/learning-from-the-cambridge-analytica-fiasco

Ng, R., & Levy, B. (2018). Pettiness: Conceptualization, measurement and cross-cultural differences. *PLOS ONE*, *13*(1). https://doi.org/10.1371/journal.pone.0191252

Ng, R., Lim, S.Q., Saw, S.Y., Francis-Tan, C., & Tan, K.B. (2018). 40-year Projections of disability and living arrangements of older adults in Singapore. Submitted for publication.

Ng, R., Hiew, Y.L., Goh, H.L., & Tan, K.B. (2018). Implementing a nationwide predictive model for frequent readmissions in Singapore public hospitals. Submitted for publication.

Ng, R., Allore, H., Trentalange, M., Monin, J., & Levy, B.R. (2015). Increasing Negativity of Age Stereotypes across 200 Years: Evidence from a Database of 400 Million Words. *PLOS One*, *10*(2), e0117086. doi: 10.1371/journal.pone.0117086.

PRWeb. (2012). Garner predicts cloud computing spending to increase by 100 percent in 2016. *PRWeb*. Retrieved from http://www.prweb.com/releases/2012/7/prweb9711167.htm

About the Author

Reuben Ng spent 16 years in government, consulting, and research. In government, he was in the Prime Minister's Office of Singapore driving evidence-based policy-making through data analytics, and Smart Nation strategies. In consulting, he co-built the Advanced Analytics practice at a top firm and implemented complex analytics projects across industries and functions. In research, he is an expert in quantitative social sciences, social gerontology, and credited with creating innovative techniques to measure societal perceptions/stereotypes that are applied to policy, and program evaluation. Prof. Ng trained as a behavioural scientist at NUS, Oxford and Yale where he was Singapore's first Fulbright Science and Technology Scholar.

Politics and Governance (ISSN: 2183-2463)

Politics and Governance is an innovative new offering to the world of online publishing in the Political Sciences. An internationally peer-reviewed open access journal, Politics and Governance publishes significant, cutting-edge and multidisciplinary research drawn from all areas of Political Science.

www.cogitatiopress.com/politicsandgovernance

COGITATIO