



Public data primacy: the changing landscape of public service delivery as big data gets bigger

Michael Overton¹ · Sarah Larson² · Lisa J. Carlson¹ · Stephen Kleinschmit³

Received: 12 October 2022 / Accepted: 16 October 2022 / Published online: 27 October 2022
© The Author(s), under exclusive licence to Institute for Global Public Policy, Fudan University 2022

Abstract

The growth and expansion of “Big Data” is fundamentally changing public service delivery. Big Data is getting “bigger,” and public organizations will have new opportunities to cultivate and challenges to address. To understand the effects of the growth of data on public organizations, we introduce the Public Data Primacy (PDP) theoretical framework, which builds on existing scholarship through four propositions about data, technology, and its use in the public sector. The framework posits that public sector work will become increasingly data-centric as data continues to get “bigger.” Ultimately, the PDP leads to two predictions about the public sector. First, we predict that the primacy of data in the delivery of public services is inevitable. Second, this forthcoming reality will require public servants to adopt new models of public service oriented around data. The PDP theoretical framework provides a systematic lens in which public administration scholarship can evaluate the future of data growth and its impacts upon public service delivery.

Keywords Big data · AI · Public service delivery

The growth of “Big Data” is fundamentally altering how public organizations make decisions and provide public services. Public employees plan and deliver public

✉ Michael Overton
moverton@uidaho.edu

Sarah Larson
sarah.larson@ucf.edu

Lisa J. Carlson
lcarl@uidaho.edu

Stephen Kleinschmit
swklein@uic.edu

¹ Department of Politics and Philosophy, University of Idaho, Moscow, ID, USA

² School of Public Administration, University of Central Florida, Orlando, FL, USA

³ Department of Public Administration, University of Illinois Chicago, Chicago, IL, USA

services using their knowledge and experience. However, the limitations of human cognition, information availability, and time (Simon, 1955) have historically constrained the ability of public employees to make optimal decisions. The foregoing limitations generate what is known as bounded rationality (Simon, 1955), wherein decision-makers select an act leading to an outcome that satisfies or is ‘good enough’ given the various constraints. In contrast, informed users of Big Data have the potential to mitigate, or perhaps even overcome, some of these constraints as shown in studies linking the proper use of Big Data to an improvement in the precision, accuracy, and optimality of decision-making processes and outcomes (Alkathiri et al., 2020; Andrews et al., 2016; Desouza & Jacob, 2017; Guirguis, 2020; Maciejewski, 2017; van der Voort et al., 2019).

Big Data is changing more than how decisions are made. There is a fundamental shift occurring in the operation of public organizations that is rooted in the use of data. Bureaucratic discretion is increasingly practiced behind a keyboard rather than on the street (Busch & Henriksen, 2018), networks of sensors are replacing networks of peers in policy processes (Rabari & Storper, 2015), and data are driving organizational decisions rather than administrative experience, knowledge, and social networks (Okwechime et al., 2018). The emergence of Big Data is having a profound impact on public organizations.

Moreover, Big Data is getting “bigger,” and with it, public organizations face new opportunities and challenges. To date, Big Data scholarship in public administration has focused on specific data applications rather than the broader trends associated with data-driven government. Less prominent in existing scholarship is an understanding of how the growing availability of data influences and alters the bases of administrative behavior. How does the growth of data affect its use in the public sector?

To understand the effects of data on public organizations, we introduce the Public Data Primacy (PDP) framework. The PDP draws from existing empirical and theoretical studies to construct a novel theoretical framework consisting of four propositions about data in the public sector. Proposition 1 establishes the concept of DATA, an abstract point of complete data saturation. Three additional propositions follow given the existence of DATA as outlined in Proposition 1. Proposition 2 is a claim about work-related data centrality, and Propositions 3 and 4 connect data qualities and data use outcomes, respectively.

The PDP is presented as the first step in constructing a rigorous and fully specified theory to begin discussions pertaining to data growth and its use in the public sector. In its current form, the PDP is a theoretical framework with informal propositions that can be formalized in future research. Therefore, the PDP is not presented as, nor should it be considered at this stage, a fully realized theory.

The PDP framework posits that public sector work becomes increasingly data-centric as Big Data gets “bigger.” More data improves outcomes, and better outcomes incentivize greater use of data-driven processes in public organizations. Since there are no indications that data will stop growing, public organizations will increasingly rely on data to deliver services. Ultimately, the PDP leads to two bold predictions about the public sector:

1. The primacy of data in the delivery of public services is inevitable.
2. Public servants require new models of public service oriented around data.

This work begins with a discussion of the transition from small data to Big Data environments. We examine the factors driving the growth of data, how Big Data is more than simply lots of data, and how applications such as Artificial Intelligence (AI) are changing the public sector. Next, we outline the four propositions of the PDP framework that uncover how the continued growth of data is likely to affect the data centrality of public organizations. We conclude with the implications of the PDP for the practice and scholarship of public administration.

1 The new datafied world

1.1 From small to big data

Data have long served an important purpose in society to record, understand, and analyze the world (Mayer-Schönberger & Cukier, 2013). At their most basic, data are discrete facts that are unaggregated, raw, and have not been converted into information (Jennex & Bartczak, 2013). More broadly, data are an abstraction of a real-world phenomenon (Kelleher & Tierney, 2018) that are typically conceptualized as the numerical depiction of measurable phenomena (Priestley & McGrath, 2019). For this paper, data are “an atomic unit that can be captured—measured, seen or heard—and thus extracted, analyzed, and converted into information and ultimately into new knowledge” (Priestley & McGrath, 2019, p. 97).

The world is becoming awash in data. The global data generated daily is an estimated 2.5 quintillion bytes (Marr, 2018), with the whole data universe estimated to grow to 175 zettabytes by 2025 (Reinsel et al., 2018). This explosion of data is a direct consequence of the increased datafication of society—quantifying individual activities and social functions into a format that can be stored and analyzed (Mayer-Schönberger & Cukier, 2013).

Lazer and Radford (2017) argue that the growing datafication can be explained through the proliferation of three data categories: digital life, digital trace, and digitalized life data. First, digital life refers to data captured on individuals’ digitally mediated social behaviors or actions across online platforms like Twitter. These data can represent both actions of a non-digital society and a realm of human behavior that exists as distinctively digital, where information and interactions can be filtered and adjusted. Second, digital trace data refers to administrative data or bureaucratic records of action taken, but it is not the action itself, unlike digital life data. For example, digital trace data includes voter registration records but not the act of voting. Finally, digitalized life refers to non-digital behavior captured and represented in a digital form. Information and Communication Technology (ICT) captures vast quantities of our non-digital activities through sensors, cameras, and phones. For example, video footage capturing an individual crossing the street from a traffic cam is considered digitalized life data.

Datafication at its current rate is made possible by three historical trends: exponentially increasing computing power, the shift from analog to digital communication technologies, and the positive feedback loop between data generation and new technologies. First, the most fundamental driver of change in the past century has been the exponential growth of computing capacity. Moore's Law suggests that general computing power would roughly double every 12–18 months due to improvements in computer hardware (Moore, 1965). The exponential growth in computing capacity has enabled ever-increasing information collection, storage, and processing capacity. As these capacities have increased, they have incentivized the creation of new platforms and data structures that can take advantage of these gains.

Computing capacity alone is not enough to explain widespread datafication. Second, the shift from analog to digital ICTs that began in the 1950s and increased dramatically in the 1990s is a fundamental reason for the explosion of data (Brady, 2019). Digitizing ICT activities has increased the datafication of communication activities once limited to face-to-face interactions while simultaneously increasing the reach of our networks and the ease of communicating with others regardless of distance (Brady, 2019). The growth of digital life (i.e., digital platforms), digital trace (i.e., administrative data), and digitalized life (i.e., datafication of real-world phenomena) data (Lazer & Radford, 2017) cannot be explained without acknowledging the role of digital ICTs.

The final factor driving data growth is the positive feedback loop between data generation and new technological development (Clauson, 2020). The data created by digital and sensor technologies is vital for improving current technology and developing the next generation of technologies, such as AI. As technology improves, the capacity to collect, store, and use data increases. Consequently, new technologies emerge to take advantage of their increased usability. For example, AI applications both create and use data to improve the function of the application (Cockburn et al., 2019).

1.2 Big data is more than lots of data

The net result of Moore's Law, the digitalization of ICTs, and the positive feedback loop between data generation and new technology is a transition from a small to a Big Data world. Small data are limited in volume, narrow in scope, expensive to acquire, "analog," and collected sporadically (Kitchin, 2014). The use of small data to inform organizational decisions and create knowledge is limited due to the relatively high cost of recording and analyzing data in an analog format. Improvements in computing power and the digitalization of ICTs have lowered the relative costs of collecting, storing, and analyzing large volumes of data enabling the emergence of Big Data. Big Data is large, multi-purpose, digitally recorded, continuously created, and relatively cheap to generate, acquire, and store (Kitchin, 2014). Private companies increasingly treat data as a fundamental commodity (Cockburn et al., 2019). Big Data is also having a profound impact on the public sector by changing how public sector organizations are managed (Mullich, 2013; Rogge et al., 2017), services are provided (Sarker et al., 2018), policy is designed (Desouza & Jacob,

2017), programs are evaluated (Gerrish, 2016), and what skills organizations will need (Klievink et al., 2017; Overton & Kleinschmit, 2021; Secundo et al., 2017).

The significance of Big Data can be understood from a technical and conceptual perspective. From a technical perspective, the properties of Big Data—three V's—volume, velocity, and variety—distinguish it from small data (Desouza & Jacob, 2017; Kitchin, 2014). Volume refers to the large amount of data that gets collected and stored. Velocity refers to the speed at which data are created and collected. Big Data is being created continuously, whereas small data is created infrequently due to relatively large time intervals between collection periods. Variety, the final V, refers to the multitude of different data types created, such as audio, visual, video, spatial, and text. Mergel et al., (2016, p. 931) define Big Data within the specific context of public affairs as “high volume data that frequently combines highly structured administrative data actively collected by public sector organizations with continuously and automatically collected structured and unstructured real-time data that are often passively created by public and private entities through their internet interactions.” From a technical perspective, the transition from small to Big Data is changing the size, speed, and types of data created and collected.

From a conceptual perspective, Big Data is the result of the datafication of society and subsequent data munificence (Mayer-Schönberger & Cukier, 2013). Compared to small data, Big Data is more exhaustive, granular, and relational (Kitchin, 2014). Exhaustivity refers to the movement away from samples and towards acquiring populations of data or “ $n = \text{all}$ ” (Mayer-Schönberger & Cukier, 2013). Granularity refers to the increasing resolution of Big Data where more refined quantifications are being datafied. Relationality refers to the ability to connect or relate data collected from different sources. The increased exhaustivity and granularity create new opportunities to connect different datasets. For example, social media data can be related to administrative data by connecting individuals' names. In total, the exhaustivity, granularity, and relationality of Big Data suggest that the world is increasingly data rich due to the comprehensiveness of datafication.

1.3 Big data and AI: changing the public sector

The technical and conceptual aspects of Big Data are having a considerable impact on public service delivery. Many scholars highlight its potential benefits, such as improved government-citizen understanding (Clarke & Margetts, 2014), better alignment between citizen preferences and government services (Chen & Zhang, 2012), improved responsiveness (Mergel et al., 2016), performance (Klievink et al., 2017), and decision-making (Desouza & Jacob, 2017). In contrast, others highlight the perils of Big Data, from its accidental or purposeful mishandling and misuse (Schintler & Kulkarni, 2014) to the ways in which it can exacerbate inequalities (Busuioc, 2020), punish the digitally invisible (Mergel et al., 2016), and decrease regulatory oversight (Sun & Medaglia, 2019). Scholarship on the potential positive and negative consequences of the use of Big Data varies widely and is somewhat speculative.

What is clear, however, is that one key area already reshaping the administration of government work is automation using AI (Maciejewski, 2017). AI is a broad concept with an array of definitions. Simply put, it is a technical system that can learn (Panch et al., 2018) and take action from information or environmental stimuli (Zuiderwijk et al., 2021). Alternatively, it can be thought of as the creation of artificial agents (Gahnberg, 2021). AI can be employed in various ways, from automating tasks to aiding in complex decisions as a supporting tool. Data are central to the success and application of AI because it uses data to inform operations through machine learning. Machine learning is a set of techniques within AI methods that “learn” from associations in data (Panch et al., 2018).

Automation from AI is changing the practice of bureaucratic discretion. When developing regulatory plans, governments and bureaucrats need to understand AI’s scope, benefits, and risks (Taeihagh, 2021). While there are many concerns about AI-informed decision-making in the public sphere, increased use from the growth of data is relevant in terms of regulation and “artificial discretion” (Young et al., 2019). AI presents a difficult challenge for regulators because of algorithmic or “black box” decision-making (Sun & Medaglia, 2019). Governmental regulators in situations where theoretical or intuitive explanations cannot be explored, like the 2008 financial crisis, search for meaning in more and more data (Kempeneer, 2021). New techniques for explaining and evaluating “black box” algorithms are being employed (Gerke et al., 2020), which helps regulate AI. However, these techniques require regulators to understand and work with the data used to “train” their algorithms. Regulating AI incentivizes greater collection and use of data in governments.

The transition from small to Big Data is having profound effects on governance. Of note, Big Data is powering AI, which is changing the role of public officials in public organizations. The following section presents the PDP theoretical framework, which outlines how the continued growth of data and AI incentivizes data use in public organizations.

2 Public data primacy (PDP) theoretical framework

By all indications, the growth of data and advancement of AI are not stopping or plateauing in the near future. Given their extensive impact on governance, it is crucial to understand how data influence public organizations. As stated at the outset, this manuscript presents a theoretical framework—the PDP—describing how the expansion of data affects their use in the public sector. In addition, this framework offers informed, initial propositions about the impact of data on public sector organizations. The PDP is presented in the next four subsections with each subsection culminating in a proposition about the likely influence of Big Data on governance.

2.1 “Bigger” data

The transition from small to Big Data is ongoing as data continues getting “bigger.” Over time, the volume, velocity, variety, exhaustivity, granularity, and relationality of data increase, resulting in more *comprehensive* data. Figure 1 illustrates this idea. In the abstract, the shift from small to Big Data can be represented as a data continuum. The smallest unit is a single datum, representing a small aspect of the natural world and providing little information. On the opposite end of the continuum is the recording of “the totality of information” when “ $n=all$ ” (Mayer-Schönberger & Cukier, 2013, p. 197). More comprehensive data will represent phenomena with greater precision, comprehensiveness, verisimilitude, and confidence.

Yet, data alone are meaningless. *Data devices* and employees with *data skills* are also required to create value from data. Data devices include analytical methods and applications that extract information from or use data. Different analytical methods help analysts understand and find connections in data that would not be possible otherwise. In addition, data applications like AI and decision support tools can help automate and augment tasks. These tools help condense vast data streams into useful, concise information. Data skills are also necessary to make sense of data and find meaningful ways to use results. Humans add value to AI through their ability to capture, curate, analyze, and apply data-informed insight to a problem (Overton & Kleinschmit, 2021; Young et al., 2019). Data skills allow individuals to ensure appropriate data are being fed into AI platforms and that any decisions resulting from AI can be evaluated and understood, creating verifiable AI (Wirtz & Müller, 2019). For example, Janssen et al. (2022) found that prior professional knowledge, such as data skills, helped governmental officials determine when machine learning algorithms provided incorrect suggestions.

Combined with improved data devices and skills, comprehensive datafication or $n=all$ can generate a 1-to-1 perfect simulation of phenomena or even reality itself. We refer to the concept of a perfect simulation of reality as *DATA* for ease of reference. *DATA* is a function of $n=all$, data devices, and data skills. Increases in the technical

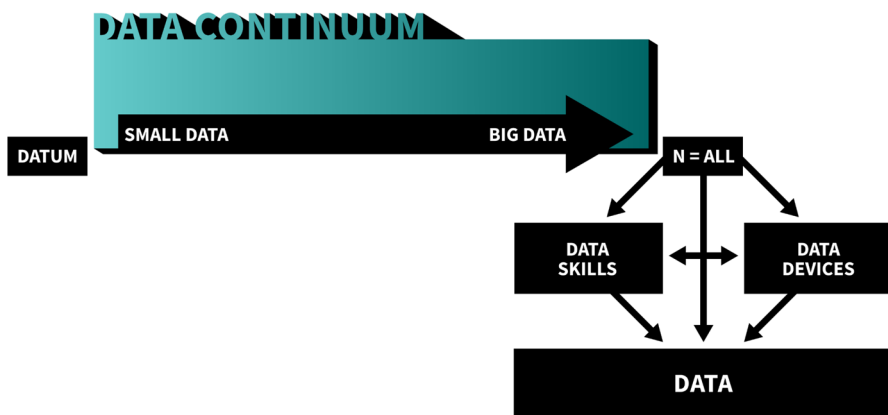


Fig. 1 Data continuum and DATA

and conceptual defining characteristics of Big Data and improvements in the sophistication of data devices and adoption of data skills move society closer to DATA. The value of data to public organizations increases as data approaches DATA. The foregoing discussion leads to the first proposition of the PDP framework:

Proposition 1: DATA is a function of the comprehensiveness of data (i.e., $n=\text{all}$), data application and analysis tools, and data skills.

2.2 Linking data growth, utilization, and outcomes

As data approaches DATA, it will yield more value for public officials and organizations. The value of using data in public organizations comes from data use outcomes. Potential improvements in data use outcomes incentivize public officials and organizations to use data in new ways. Once used in new ways, the perceived and actual improvements incentivize continued use and increase the probability it will be used in new ways to deliver public services. In short, as data gets bigger (i.e., closer to DATA), data use outcomes improve, which incentivizes organizations to become more data-centric, as shown in Fig. 2.

Data centrality refers to the breadth and depth of the integration of data use in the public sector. The breadth of data centrality refers to “who” is using data in public organizations and is determined by individuals’ and departments’ functional needs and task goals. For individuals, data are not in the sole domain of senior administrators and department heads making strategic decisions and will increasingly become helpful to front-line employees for operational purposes. Already, Big Data is changing the skills needed to govern effectively (Overton & Kleinschmit, 2021) and altering how public employees practice bureaucratic discretion (Busch & Henriksen, 2018).

For organizations, multiple departments within a government bureaucracy orient their use of data around the department’s mission and goals. The substantially different data practices and varying levels of integration across departments are anchored in their functional needs (Berardo & Lubell, 2016). While budgeting departments have been using data extensively to record how public funds are spent for the last 100 years, they have been used in new ways (e.g., planning) to improve budgeting practices over time (Schick, 1966). Conversely, communication departments are only recently integrating extensive data analysis into their operations because of the recent proliferation of social media analytics (Belkahla Driss et al., 2019).

The depth of data centrality refers to “how” data are applied to public services, which typically occurs in decision-making. For example, performance management systems use performance measures to guide budgeting decisions, inform planning, and evaluate an aspect of a program. However, data can be used for more than informing decisions and can be directly used to provide services. Any AI, automated, or digital



Fig. 2 Data centrality

service requires data to train or build the service and a stream of input data so it can create a specified output.

There are also challenges associated with Big Data and AI that could undermine data use outcomes and decrease data centrality, such as equity, representation of the digitally invisible, and algorithmic bias. However, *more data*, not less, can solve these issues. One approach to addressing algorithmic bias includes collecting relevant demographic data and using that data to correct for identified bias (Barocas 2017)—a solution grounded in acquiring and using more data. In the case of the digitally invisible, individuals experiencing homelessness were given temperature detecting devices during a heat wave to understand their service needs (Longo et al., 2017). More data, not less, were required to serve the digitally invisible better.

Therefore, these challenges further incentivize data centrality. Data can improve equity in automated processes, identify and address algorithmic bias, and quantify those currently unrepresented in data. At some point in the future, the growth of data, data devices, and data skills ensures that the best way to address these issues is with greater data centrality. While the challenges of using Big Data and AI to deliver public services are real, they do not undermine data use outcomes. The second proposition of the PDP framework is:

Proposition 2: As data approaches DATA, public sector work will become more data-centric.

2.3 Why “bigger” data matters

The transition from small to Big Data elevated three important data qualities—convenience, instrumentality, and authority—that increase as data grows. These three qualities are important because they are the underlying factors that establish the range of data use outcomes and the potential degree of data centrality in public organizations. They enable and constrain the utility derived from data use outcomes and the possible depth and breadth of data centrality in public organizations.

The growth of data has improved the *convenience* of collecting data because the relative cost of capturing, curating, and applying data has decreased (Kitchin, 2014). As data approaches DATA, data’s convenience will minimize barriers preventing its use, which limits an organization’s data centrality. Data are an information good (Mihet & Philippon, 2019), and once generated, they can be used multiple times relatively cost-free. In addition, data on digital activities are cheap to create, find, and can be automatically generated and collected (Goldfarb & Tucker, 2019).

However, data collection is not costless, though costs will continue to decrease. Capturing data requires relatively expensive infrastructure investment. Yet, once in place, the actual cost of collection is low, resulting in a considerable economy of scale for large collection efforts (Haskel & Westlake, 2018). The infrastructure required for data collection and storage will get cheaper as the need for it becomes more ubiquitous across public organizations. The initial infrastructure costs are also likely offset by the substantial savings in operational costs (Maciejewski, 2017).

As data approaches DATA, so will data’s *instrumentality*, which refers to the ability to apply data to public sector problems and achieve desired outcomes. The

instrumentality of data enables and constrains data use outcomes and an organization's data centrality in two specific ways: greater precision and accuracy from analysis and broader potential applications. First, information theory and the information processing view of organizations (Galbraith, 1974) suggest that Big Data can be applied more precisely and accurately to a broader array of problems than small data (Brynjolfsson et al., 2011).

Unlike manual collection, data streams allow real-time assessment and response, reducing delay and enhancing precision. It becomes increasingly possible to create comprehensive, accurate, and precise models that are less probabilistic and more deterministic. "Certainty through saturation" is when no new information is derived through data collection (Francis et al., 2010), where confidence intervals narrow to negligible sums, and confidence levels approach 100%. Phenomena can be modeled with near-total certainty rather than a "satisficed" model grounded in the limitations of prior manual collection and reporting methods.

Second, the growth of data broadens its potential applications in the public sector. The comprehensiveness of data enables and constrains the set of public sector problems that data can address (Williamson, 2014). The public sector's wholesale adoption of data-driven practices has been limited by its ability to quantify policy problems and solutions (Lindblom, 1959). As data approaches DATA, more phenomena will be quantified, which increases the possible ways it can be used in public organizations. Greater data coverage and granularity increase the likelihood that (1) data can be acquired to address specific or unique public problems and (2) a multitude of different types of data required to address complex, wicked problems can be collected. In short, the ability to apply data to a broader array of problems will increase potential data use outcomes and the maximum potential data centrality of a public organization.

Data are becoming an *authority*—a trusted source of knowledge. Once produced, government work becomes viewed in comparison with that data (Meijer, 2018; Meijer & Thaens, 2018). More broadly, the ability to capture, store, analyze, and visualize massive amounts of data has created an epistemological shift in science, resulting in a data-intensive empiricism (Kitchin, 2014). This epistemological shift is spilling over into the practice of public administration. Rather than relying solely on knowledge informed by theory and prior experiences, data are increasingly required to inform decisions and the delivery of public services (Rabari & Storper, 2015).

Meijer (2018) argues that new technologies have an interactive effect on the social structure of a community and that the growth of data from smart city technologies changes individual and organizational incentives because new technologies, such as Big Data, alter the perceptions and interactions of individuals and organizations with society (Orlikowski, 1992). Kempeneer (2021, p. 1) refers to this phenomenon as the "big data state of mind," which is "the state of mind that one can or should rely on large data sets rather than theory to produce valid knowledge claims." While we believe dismissing the role of theory in the decision-making process is a bridge too far, there is little doubt that the abundance of Big Data is fundamentally changing the role of data in knowing and knowledge production (Meijer, 2018). The discussion above leads to the third proposition of the PDP framework and is as follows:

Proposition 3: As data approaches DATA, the convenience, instrumentality, and authority of data will improve data use outcomes.

2.4 Data use outcomes

As data approaches DATA, the convenience, instrumentality, and authority of data increase, which increases the potential, perceived, and actual benefits derived from data use outcomes. Drawing from existing public administration scholarship, five data use outcomes are identified and explained below.

2.4.1 Outcome 1: better public sector performance at a lower cost

Public managers are motivated to improve organizational performance (Meier et al., 2015), and Big Data facilitates improved performance by increasing organizational capabilities (Andrews et al., 2016; Guirguis, 2020). Public administrators are already improving the performance of public organizations using Big Data through better service delivery, regulatory capacity, internal management practices, task automation, and decision-making (Maciejewski, 2017). Alkatheeri et al. (2020) found that the quantity and quality of Big Data improved the quality of decision-making in Abu Dhabi Governmental Organizations. Data's instrumentality improves decision-making and service delivery in the public sector, and data's convenience will help produce improvements in performance and at increasingly lower costs. The subsequent improvements in the quality of decisions and public services at lower costs will incentivize public managers to use data.

Big Data improves the precision and accuracy of decisions in public organizations, transforming the boundedly rational processes of public managers into instrumentally rational processes (van der Voort et al., 2019). The limitations that generate bounded rationality, (i.e., human cognition, information availability, and adequate time) are mitigated, and potentially overcome, using data. Overcoming these limitations leads to improvements in the precision, accuracy, and optimality of decision-making processes and their outcomes. In practice, data-driven decision-making—making decisions based on analysis rather than intuition (Brynjolfsson et al., 2011)—is not new, as governments have used data to make decisions for a long time (Schick, 1966). However, data-driven decision-making represents an improvement in public organizations' decision efficacy compared to traditional, boundedly rational decision-making processes (Hwang et al., 2021).

For example, large data techniques now allow practitioners to engage in latent dimension identification (Qin et al., 2020), which creates more information to support public services and decisions (van der Voort et al., 2019). Under previous technical paradigms, the limitations of computational power and data availability have placed much of the purposiveness of the analysis on the researcher, beset with the limitations of their cognitive capacity and biases. Rather than engaging in discovery through a priori theoretically initiated lines of inquiry, large data techniques now allow practitioners to engage in latent dimension identification. As such, machine

learning has the potential to lead to a more agnostic approach in that an examination of all potential relationships within data gives a deeper understanding of the phenomena. As the practice of administration becomes less deductive and driven by existing theory, it will move towards a more abductive state of practice where important relationships are revealed, and meaning can be interpolated through large data-sets (Haig, 2020).

Separately, the growth of AI within public service is accelerating performance improvements from data. For example, AI has successfully forecasted high crime risk transportation areas, which increased both the quality of public transportation and the public's overall safety (Kouziokas, 2017). AI-enabled IT platforms improve the performance of capital planning and budgeting processes, which lead to better budget preparation and implementation (Wang 2022). The growth of smart cities has led to the use of AI to better collect, manage, and analyze data, which is leading to faster and smarter systems for public service delivery (Allam & Dhunny, 2019). Thus, advanced data technologies can provide an important mechanism for addressing complex problems through enhanced capacity for evaluation. As data approaches DATA, performance gains from data use incentivize public managers to find novel ways to use data in public organizations. Opportunities to use new data and data applications to improve an organization's performance emerge as data grows. Public managers will increasingly seek innovative ways of applying new data in search of novel performance gains. More granular or comprehensive data can be applied to situations where data are already being used, further integrating data into public sector work. The potential for further performance improvements at lower costs could exponentially increase the data centrality of public organizations.

2.4.2 Outcome 2: process legitimacy

Public organizations seek political and social legitimacy to gain acceptance from critical stakeholders, peer institutions, or the public at large (Aldrich & Fiol, 1994). An organization's legitimacy involves a general perception that the process of determining actions taken is appropriate and desirable within societal norms (Suchman, 1995). The rise of open data introduced a major digital mechanism for ensuring government legitimacy, offering access to public data as a means to improve transparency and accountability (Attard et al., 2015). As data approaches DATA, public institutions use data to increase the legitimacy of the process by which decisions are made. Government stakeholder perceptions that data are required for the legitimacy of public services continually incentivize public officials to use data to gain legitimacy.

The authority and the instrumentality of data will increase the perceived legitimacy of the actions taken by a public organization. Data increasingly becomes a precondition of any legitimate and acceptable process or conclusion as data becomes more authoritative. Integrating data into public sector work also provides a path toward transparency in the decision-making process, which improves citizen trust (Altayar, 2018), worker engagement, and subsequently perceived legitimacy of a government institution from the public and employees (Desouza & Bhagwatwar,

2012). Data's instrumentality in data-driven decisions is also associated with more accurate decision-making (Brynjolfsson et al., 2011), which can be crucial for building public trust and legitimacy (Holden & Organization for Economic Co-operation and Development, 2015). Improved performance from data use further entrenches the view that legitimate decision processes require data.

2.4.3 Outcome 3: isomorphic legitimacy

Organizations conform to the professional practices of peer institutions over time to gain institutional legitimacy, resulting in homogeneous structuration from institutional isomorphism (DiMaggio & Powell, 1983). Public organizations are particularly vulnerable to isomorphic pressures (Frumkin & Galaskiewicz, 2004). As data approaches DATA, more institutions will use data to improve data use outcomes, and consequently, the marginal isomorphic benefits of its use will increase. Public leaders have a wide range of data perceptions ranging from technophobia to enthusiast (Guenduez et al., 2020). Even reluctant public officials are likely to be enticed to use data because the public sector is headed toward widespread adoption of data-centric practices. The application of data in public organizations is a self-reinforcing activity where data use outcomes increase the data centrality of public institutions, increasing the value of data to gain institutional legitimacy. As more institutions use data, public managers are likely to "herd" around data use to gain institutional legitimacy.

2.4.4 Outcome 4: mitigates organizational risk and the risk of undesirable policy outcomes

Public officials are risk-averse (Nicholson-Crotty et al., 2019), and the instrumentality of data can be used to mitigate certain risks. Unfortunately, all public decisions and services come with some level of political risk. The instrumentality of data suggests that it can be employed to reduce the risk of undesirable policy outcomes associated with decision-making and the provision of social services.

There are two significant issues that public servants face, each of which involves decision-making and the attempt to manage risk. First, to choose a public policy is to choose a gamble, with risk understood here to mean the probability of realizing a downside payoff for a failed policy outcome. As data approaches DATA, the risk (probability) of realizing a downside payoff from failed service outcomes decreases. There are at least two reasons why the probability decreases. First, increased utilization of data facilitates more precise and purposeful government processes. Second, public officials explicitly adopt and formalize routines and decision rules to mitigate risk (Wolman & Spitzley, 1996). With the increased instrumentality of data, routines and decision rules can be formalized and quantified with extreme precision and consistency. Consequently, instrumentality decreases the risk of making decisions that might lead to a Pareto-suboptimal provision of public services. Simply put, public sector reliance on data reduces the risks associated with public policy failure.

2.4.5 Outcome 5: mitigating individual risk: data as a scapegoat

If the decision-maker chooses a policy that leads to policy failure, then a second and related risk for an individual public service employee is realized in the form of electoral consequences (Mayhew, 2004). An individual's ability to bear the risk associated with policy failure is a critical component of public entrepreneurship—i.e., innovative decision-making and program delivery in high-risk conditions (Schneider et al., 1995). Moreover, it has been reported that delegating responsibility is a key tactic in avoiding or shifting blame away from the individual (Epstein & O'Halloran, 1999).

Data provide many benefits to public officials wishing to avoid political and bureaucratic consequences by shifting the responsibility of negative consequences from the individual or organization onto the data itself. The primary reason is that public officials are shielded from electoral or bureaucratic consequences for inferior outcomes because the public perceives the reliance on and use of data as authoritative. Therefore, data will reduce, but not completely eliminate, the electoral and bureaucratic risks associated with decision-making in government. In the end, the five outcomes associated with data use described in detail above produce a fourth proposition of the PDP framework:

Proposition 4: As data approaches DATA, data use outcomes—potential, perceived, and actual—will incentivize increased data centrality in the public sector.

To summarize, the PDP is a theoretical framework that generates four key propositions connecting the growth in data to the increasing reliance on data in public organizations. Proposition 1 establishes the concept of DATA, an abstract point of complete data saturation representing what Big Data is moving towards. Proposition 2 outlines a causal argument connecting DATA, data use outcomes, and data centrality in public organizations. Proposition 3 explains how the growth of data improves data use outcomes via an increase in data convenience, instrumentality, and authority. Finally, Proposition 4 highlights five data use outcomes that incentivize the growth of data centrality.

3 Conclusion and future implications

As scholars, we face a daunting challenge: studying and understanding an emerging data-driven reality. The PDP theoretical framework provides a foundation for understanding how data and AI change the public sector. The need to understand how data are changing public organizations and develop practitioners for an increasingly data-centric public sector will only increase. Through careful study and thoughtful guidance, we can prepare public administrators for the primacy of data in the public sector.

The PDP theoretical framework leads to the expectation that public officials and organizations will be increasingly incentivized to become more data-centric because of the potential of data use outcomes' perceived and realized benefits. This

framework and its four key propositions are based on the underlying assumption that Big Data will continue to grow in volume, variety, velocity, exhaustivity, granularity, and relationality. The pinnacle of this process is a near-perfect simulation of reality, referred to as DATA. Moreover, increases in the comprehensiveness of data collected, data devices, and data skills will increase the convenience, instrumentality, and authority of data use.

The depth and breadth of data's integration into public service can demonstrate different types and degrees of data centrality and presents an important opportunity for scholarly inquiry. The specifics of the changes in the public sector's data centrality as society moves toward DATA are purposefully left ambiguous as the exact changes and ordering of the changes are currently speculative. If correct, then the short- and long-term implications of an increasingly data-centric public sector are significant for the practice of public administration. The two most important predictions emerging from the PDP are that (1) the primacy of data in the public sector is inevitable, and (2) public administrators will need to become public data servants.

Prediction 1: The primacy of data in the delivery of public services is inevitable

The PDP theoretical framework suggests that data will become irreplaceable to public organizations establishing the primacy of data in delivering public services. The growth of data means that society will continue to approach DATA for the foreseeable future. Consequently, the benefits derived from data-driven outcomes will continue to increase for the foreseeable future, incentivizing increased data centrality. For these reasons, the primacy of data in the provision of public services is only a matter of time. Data will eventually be *an* essential, if not *the* key, resource required to deliver public services and conduct work in the public sector.

Prediction 2: Public servants require new models of public service oriented around data

The PDP theoretical framework not only suggests that data primacy is inevitable but that public administrators will need to reframe their public service ethos toward an increasingly data-centric public sector. Public administrators must become, at least partially, public data servants. A public data servant is more than a database manager or an analyst in a public organization. Instead, they understand the importance of data in a democratic society and work to ensure data are used fairly, transparently, equitably, and appropriately. The Covid-19 pandemic provided examples of how data might reshape the value administrators add to public service delivery, such as protecting data from political manipulation (Luscombe, 2021) or presenting data fairly (Engledowl & Weiland, 2021). Beyond the issues mentioned above, public data servants will need to address the ethics and values of data use. The ethical issues associated with Big Data and AI, such as privacy, will require specialized data devices and skills that enable public servants to identify and address concerns as they emerge.

The PDP theoretical framework provides scholars with numerous avenues to conduct future research formalizing and empirically testing the propositions presented in this manuscript. Research based on the framework's four propositions can conceptually refine, and empirically measure important data concepts of interest in public administration scholarship. For example, another data-centricity consideration that should be explored by future research is the application of data at different

phases of the policy cycle—problem identification, agenda-setting, policy formulation, implementation, and evaluation. Each phase of the process is associated with different actors, stakeholders, and institutional incentives, which change the benefits derived from data use outcomes. Understanding the data centrality of the policy cycle presents another opportunity to apply the PDP and understand data's integration in governance.

Future research should also consider the needs of practitioners in a data-centric public sector. As data becomes more central to the work of public organizations and the delivery of public services, public administrators will require public service models that understand both the role of administrators and the new demands brought about by the primacy of data in public service. It is vital to monitor the needs of public administrators and help them adapt to the new, datafied world. Regardless of the specific focus of future research, these future works will provide grounds for applying the theoretical framework to the changing landscape of public service with the expansion of Big Data and AI and provide necessary guidance to the delivery of public service.

Acknowledgements We would like to thank the reviewers for their helpful comments, the guest editors of this special issue call for all the work they have done to make this possible, and a special thank you goes to Dr. Eric Stokan who provided useful feedback on an early version of the manuscript.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

- Alkatheeri, Y., Ameen, A., Isaac, O., Nusari, M., Duraisamy, B., & Khalifa, G. S. A. (2020). The effect of big data on the quality of decision-making in Abu Dhabi government organisations. In N. Sharma, A. Chakrabarti, & V. E. Balas (Eds.), *data management, analytics and innovation* (Vol. 1016, pp. 231–248). Springer. https://doi.org/10.1007/978-981-13-9364-8_18
- Allam, Z., & Dhunny, Z. A. (2019). On big data, artificial intelligence and smart cities. *Cities*, 89, 80–91. <https://doi.org/10.1016/j.cities.2019.01.032>
- Altayar, M. S. (2018). Motivations for open data adoption: An institutional theory perspective. *Government Information Quarterly*, 35(4), 633–643. <https://doi.org/10.1016/j.giq.2018.09.006>
- Andrews, R., Beynon, M. J., & McDermott, A. M. (2016). Organizational capability in the public sector: A configurational approach. *Journal of Public Administration Research and Theory*, 26(2), 239–258. <https://doi.org/10.1093/jopart/muv005>
- Brady, H. E. (2019). The challenge of big data and data science. *Annual Review of Political Science*, 22(1), 297–323. <https://doi.org/10.1146/annurev-polisci-090216-023229>
- Busch, P. A., & Henriksen, H. Z. (2018). Digital discretion: A systematic literature review of ICT and street-level discretion. *Information Polity*, 23(1), 3–28. <https://doi.org/10.3233/IP-170050>
- Busuioac, M. (2020). Accountable Artificial Intelligence: Holding Algorithms to Account. *Public Administration Review, Early View*. <https://doi.org/10.1111/puar.13293>
- Chen, Y.-C., & Zhang, J.-C. (2012). Citizen-centric e-government performance: Satisfaction with e-information. *Electronic Government, an International Journal*, 9(4), 388–402. <https://doi.org/10.1504/EG.2012.049726>
- Clarke, A., & Margetts, H. (2014). Governments and citizens getting to know each other? Open, closed, and big data in public management reform. *Policy & Internet*, 6(4), 393–417. <https://doi.org/10.1002/1944-2866.POI377>

- Clauson, M. (2020). The future of work. In B. J. Hoffman, M. K. Shoss, & L. A. Wegman (Eds.), *The Cambridge handbook of the changing nature of work* (1st ed., pp. 555–582). Cambridge University Press. <https://doi.org/10.1017/9781108278034.026>
- Cockburn, I. M., Henderson, R., & Stern, S. (2019). 4. The impact of artificial intelligence on innovation: An exploratory analysis. *The economics of artificial intelligence* (pp. 115–148). University of Chicago Press. <https://doi.org/10.7208/9780226613475-006>
- Desouza, K. C., & Jacob, B. (2017). Big data in the public sector: Lessons for practitioners and scholars. *Administration & Society*, 49(7), 1043–1064. <https://doi.org/10.1177/0095399714555751>
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147. <https://doi.org/10.2307/2095101>
- Engledow, C., & Weiland, T. (2021). Data (Mis)representation and COVID-19: Leveraging misleading data visualizations for developing statistical literacy across grades 6–16. *Journal of Statistics and Data Science Education*, 29(2), 160–164. <https://doi.org/10.1080/26939169.2021.1915215>
- Epstein, D., & O'Halloran, S. (1999). *Delegating powers: A transaction cost politics approach to policy making under separate powers*. Cambridge University Press.
- Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. P., & Grimshaw, J. M. (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology & Health*, 25(10), 1229–1245. <https://doi.org/10.1080/08870440903194015>
- Frumkin, P., & Galaskiewicz, J. (2004). Institutional isomorphism and public sector organizations. *Journal of Public Administration Research and Theory*, 14(3), 283–307. <https://doi.org/10.1093/jopart/muh028>
- Gahnberg, C. (2021). What rules? Framing the governance of artificial agency. *Policy and Society*, 40(2), 194–210. <https://doi.org/10.1080/14494035.2021.1929729>
- Gerke, S., Babic, B., Evgeniou, T., & Cohen, I. G. (2020). The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. *NPJ Digital Medicine*, 3(1), 1–4. <https://doi.org/10.1038/s41746-020-0262-2>
- Gerrish, E. (2016). The impact of performance management on performance in public organizations: A meta-analysis. *Public Administration Review*, 76(1), 48–66. <https://doi.org/10.1111/puar.12433>
- Guirguis, K. (2020). From big data to big performance—Exploring the potential of big data for enhancing public organizations' performance: A systematic literature review. *Jahrbuch Der Schweizerischen Verwaltungswissenschaften*, 11(1), 55–65. <https://doi.org/10.5334/ssas.140>
- Haig, B. D. (2020). Big data science: A philosophy of science perspective. *Big data in psychological research* (pp. 15–33). American Psychological Association. <https://doi.org/10.1037/0000193-002>
- Janssen, M., Hartog, M., Matheus, R., Yi Ding, A., & Kuk, G. (2022). Will algorithms blind people? The effect of explainable AI and decision-makers' experience on AI-supported decision-making in government. *Social Science Computer Review*, 40(2), 478–493. <https://doi.org/10.1177/0894439320980118>
- Kempeneer, S. (2021). A big data state of mind: Epistemological challenges to accountability and transparency in data-driven regulation. *Government Information Quarterly*, 38(3), 101578. <https://doi.org/10.1016/j.giq.2021.101578>
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures & their consequences*. SAGE. <https://doi.org/10.4135/9781473909472>
- Klievink, B., Romijn, B.-J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. *Information Systems Frontiers*, 19(2), 267–283. <https://doi.org/10.1007/s10796-016-9686-2>
- Kouziokas, G. N. (2017). The application of artificial intelligence in public administration for forecasting high crime risk transportation areas in urban environment. *Transportation Research Procedia*, 24, 467–473. <https://doi.org/10.1016/j.trpro.2017.05.083>
- Lazer, D., & Radford, J. (2017). Data ex machina: introduction to big data. *Annual Review of Sociology*, 43(1), 19–39. <https://doi.org/10.1146/annurev-soc-060116-053457>
- 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 implications of the digitally invisible. *Policy & Internet*, 9(1), 76–108. <https://doi.org/10.1002/poi3.144>
- Luscombe, R. (2021). Florida analyst who clashed with governor over Covid data faces arrest. *The Guardian*. <https://www.theguardian.com/us-news/2021/jan/17/florida-rebekah-jones-covid-data-analyst-arrest-warrant>

- Maciejewski, M. (2017). To do more, better, faster and more cheaply: Using big data in public administration. *International Review of Administrative Sciences*, 83(1_suppl), 120–135. <https://doi.org/10.1177/0020852316640058>
- Marr, B. (2018). *How much data do we create every day? The mind-blowing stats everyone should read*. Forbes. <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/>
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
- Meier, K. J., Favero, N., & Zhu, L. (2015). Performance gaps and managerial decisions: A Bayesian decision theory of managerial action. *Journal of Public Administration Research and Theory*, 25(4), 1221–1246. <https://doi.org/10.1093/jopart/mtu054>
- Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big data in public affairs. *Public Administration Review*, 76(6), 928–937. <https://doi.org/10.1111/puar.12625>
- Moore, G. (1965). Cramming more components onto integrated circuits. *Electronics*, 38(8), 51–58.
- Mullich, J. (2013). *Closing the big data gap in public sector*. Bloomberg Businessweek: Research Services, New York.
- Okwechime, E., Duncan, P., & Edgar, D. (2018). Big data and smart cities: A public sector organizational learning perspective. *Information Systems and E-Business Management*, 16(3), 601–625. <https://doi.org/10.1007/s10257-017-0344-0>
- Overton, M., & Kleinschmit, S. (2021). Data science literacy: Toward a philosophy of accessible and adaptable data science skill development in public administration programs. *Teaching Public Administration*. <https://doi.org/10.1177/01447394211004990>
- Panch, T., Szolovits, P., & Atun, R. (2018). Artificial intelligence, machine learning and health systems. *Journal of Global Health*, 8(2), 020303. <https://doi.org/10.7189/jogh.08.020303>
- Priestley, J. L., & McGrath, R. J. (2019). The evolution of data science: A new mode of knowledge production. *International Journal of Knowledge Management*, 15(2), 97–109. <https://doi.org/10.4018/IJKM.2019040106>
- Qin, S. J., Dong, Y., Zhu, Q., Wang, J., & Liu, Q. (2020). Bridging systems theory and data science: A unifying review of dynamic latent variable analytics and process monitoring. *Annual Reviews in Control*, 50, 29–48. <https://doi.org/10.1016/j.arcontrol.2020.09.004>
- Rabari, C., & Storper, M. (2015). The digital skin of cities: Urban theory and research in the age of the sensed and metered city, ubiquitous computing and big data. *Cambridge Journal of Regions, Economy and Society*, 8(1), 27–42. <https://doi.org/10.1093/cjres/rsu021>
- Reinsel, D., Gantz, J., & Rydning, J. (2018). *the digitization of the world from edge to core* (Data Age 2025, p. 28). International Data Corporation.
- Rogge, N., Agasisti, T., & De Witte, K. (2017). Big data and the measurement of public organizations' performance and efficiency: The state-of-the-art. *Public Policy and Administration*, 32(4), 263–281. <https://doi.org/10.1177/0952076716687355>
- Sarker, M. N. I., Wu, M., & Hossin, M. A. (2018). Smart governance through big data: Digital transformation of public agencies. *International Conference on Artificial Intelligence and Big Data (ICAIBD)*, 2018, 62–70. <https://doi.org/10.1109/ICAIBD.2018.8396168>
- Schintler, L. A., & Kulkarni, R. (2014). Big data for policy analysis: The good, the bad, and the ugly: Big data for policy. *Review of Policy Research*, 31(4), 343–348. <https://doi.org/10.1111/ropr.12079>
- Secundo, G., Del Vecchio, P., Dumay, J., & Passiante, G. (2017). Intellectual capital in the age of big data: Establishing a research agenda. *Journal of Intellectual Capital*, 18(2), 242–261. <https://doi.org/10.1108/JIC-10-2016-0097>
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99–118. <https://doi.org/10.2307/1884852>
- Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of artificial intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, 36(2), 368–383. <https://doi.org/10.1016/j.giq.2018.09.008>
- Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and Society*, 40(2), 137–157. <https://doi.org/10.1080/14494035.2021.1928377>
- van der Voort, H. G., Klievink, A. J., Arnaboldi, M., & Meijer, A. J. (2019). Rationality and politics of algorithms. Will the promise of big data survive the dynamics of public decision making? *Government Information Quarterly*, 36(1), 27–38. <https://doi.org/10.1016/j.giq.2018.10.011>

- Williamson, A. (2014). Big data and the implications for government. *Legal Information Management*, 14(4), 253–257. <https://doi.org/10.1017/S1472669614000553>
- Wirtz, B. W., & Müller, W. M. (2019). An integrated artificial intelligence framework for public management. *Public Management Review*, 21(7), 1076–1100. <https://doi.org/10.1080/14719037.2018.1549268>
- Wolman, H., & Spitzley, D. (1996). The politics of local economic development. *Economic Development Quarterly*, 10(2), 115–150. <https://doi.org/10.1177/089124249601000201>
- Zuiderwijk, A., Chen, Y.-C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*, 38(3), 101577. <https://doi.org/10.1016/j.giq.2021.101577>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.