Camelia GRĂDINARU[*]

The Big Data Effect: The Quest for a New Understanding of the Public Sphere

Abstract: This paper focuses on some complex implications of the use of Big Data: the epistemological changes from causality to correlation, from searching for reasons to finding trends, from narratives to databases. These challenges are applied to the public sector for a better understanding of their intricacies. The various initiatives and directives implemented by governments in many countries have shown the widespread interest in this valuable resource, but legal and ethical regulations are still needed to establish a healthy basis for using Big Data. Also, there is a gap between the promises of Big Data for the public sphere and its actual use in public organizations around the world. At the same time, new forms of divides raise essential questions about participation and representativity.

Keywords: Big Data, epistemological challenges, database, citizen's voice, narratives.

1. Introduction

Excluding the digital divide cases, the increasing convergence of new media has effects on the public sector, the most visible of which are found in political knowledge and grassroots organizations (Snow Bailard 2017, 248). Many directives for public sector data resources are aimed at regulating the information produced by public entities (The White House 2012, Australian Government 2013, The Government of Japan 2013, The European Parliament and the Council of the European Union 2019). They provide the legal framework for public sphere information, stimulating transparency, free flow of data, and fair competition. Open data is part of public sector Big Data, depicting data that can be freely used, re-used and shared. These policies encourage the availability of data, not just for economic and business stakeholders, but "primarily for the public" to gain an increased sense of social engagement and civic participation. Even if Big Data has tremendous potential, we must keep in mind that "Big Data technologies alone are not, however, a silver bullet for transforming the public sector" (Liu 2012, 6). Nevertheless, there is a gap between the immense potential of Big Data for the public sector and its actual use: for instance, the practitioners have a predilection for using digital media merely

[*] Lecturer, PhD, Department of Communication Sciences and Public Relation, "Alexandru Ioan Cuza" University of Iași. Email: camelia.gradinaru@uaic.ro.

as a new channel of communication rather than a huge reservoir of data that should be analysed to improve public policies. Within the process of professionalization of policymakers, the subsequent curricula must be adapted to this new reality. Data analysis methods such as analytics must become a part of the education of Big Data managers, alongside prediction markets or sentiment analysis (Mergel 2016; Hu 2018). The public sector now must include both non-technical and technical requirements (Munné 2016). At least as important are the policymakers' level of literacy as well as their judgment, integrity, and ethics. The COVID-19 pandemic strongly revealed the necessity of accurate and open data in crisis situations: the citizens want official information, while many countries intensified their efforts to make the information more comprehensible and measurable. Big Data comes with many challenges (Boyd and Crawford 2012): context remains critical, accountability is important, the risk of seeing patterns where they do not exist is significant, and, not least, unequal access to Big Data could create "new digital divides". "Data-driven science" could replace the "knowledge-driven science", disrupting the classic epistemological positions and requesting "the development of a situated, reflexive and contextually nuanced epistemology" (Kitchin 2014). Also, the ethical grounds are dynamic, and we must be aware all the time that "even anonymous, public data sets can produce harms depending on how they are used" (Metcalf and Crawford, 2016). In this respect, this paper investigates some nuanced implications of the use of Big Data from its epistemological consequences to the participatory issues.

2. Public Affairs in a Big Data Environment

Big Data is a fuzzy concept (Cunningham and Thissen 2014), hard to define and often misunderstood. It is frequently correlated with the expression "wisdom of crowds" (Surowiecki 2005) to highlight the possibilities of taking advantage of the particular "collective intelligence" of the Internet. Big Data is characterized by volume, velocity, variety, and complexity (Fosso Wamba et al. 2015; Desouza and Jacob 2017, 1045), but it is also "noisy" and difficult to analyse. While information grows exponentially in various domains, industry, governments, and researchers are interested in interdisciplinary collaborations.

After the implementation of Big Data in the private sector and industry, the public sector was attracted to this asset. Nevertheless, the literature review shows that there is a substantial gap between the promises of Big Data for public affairs (Chen and Hsieh 2014) and its actual implementation in public organizations (Desouza and Jacob 2017). Even if Big Data "holds tremendous potential for policy analysis" (Schintler and Kulkarni 2014, 347) and could lead to more informed policymaking, better decisions, and greater

transparency and efficiency, "government organisations seem to still be in an orientation or contemplation phase regarding Big Data" (Klievink et al. 2017, 268), in a state of "infancy" (Desouza and Jacob 2017, 1044), or just at a "programatic level" (Desouza and Jacob 2017, 1052). Thus, the public sector is falling behind in this matter (Mullich 2013). Although a consistent part of the literature is paying attention to the ways in which Big Data can improve public affairs, there are still unused data in the public sector. There is a consistent interest in Big Data's potential, but this sector has not used data mining technologies frequently, so "there is no broad implementation of big data in the public sector" (Munné 2016, 196). Critical data studies are generally in progress and the research is still low, as Kempeneer (2021) concluded: "Despite the prominence of big data in society, its use in the public sector remains grossly understudied".

From the theoretical point of view, the definitions of Big Data are scarce; the articles focused merely on characteristics, insights, applications, and challenges for the public sector. Thus, "defining Big Data is not a popular topic in current research" (Fredriksson et al. 2017, 45). The same conclusion is drawn by Mergel et al. (2016, 929-930) as they systematize definitions of Big Data across disciplines, observing that one exception within public affairs is a White House report (The White House 2014, 3). The focus on the scale of new emerging data could obscure other significant points such as the nature of the data collected, their form (structured, unstructured, semi-structured), their source, or the absence of a context in which they could be reasonably comprehended. There is also frequently a lag between the act of collecting data and effective analysis.

Public decision-making is the main sector that has seen improvement with data analytics (Fredriksson et al. 2017, 52). Social media and open data will represent important drivers for the public sector. Governments produce and collect huge quantities of information (through taxes, the health system, traffic data, official documents); at the same time, user-generated content is significantly growing (on social networking sites, blogs, forums). The participatory citizens (Liu and Yuan 2015) are more involved in the life of their city and generally use social media as a megaphone for their opinions. The online presence of public institutions, from live streaming to multimedia posts, generates feedback from various audiences. Within a certain ethical and legal frame, these data offer access to a plethora of people's desires, choices, sentiments, or even whims. Near real-time data could be analysed now, with positive effects at the level of policies (Janssen et al. 2017). Mergel et al. (2016, 931) highlighted that in public-affairs research we are dealing with "multimodal digital data generated by public and private providers": data automatically collected by public entities, social media data, data recorded by sensors. We also should notice the significant shift in the understanding of what *public* means today and the concerns

about privacy. We also have to mention the Internet of Things (IoT) Big Data that are produced by the smart devices that are connected to the Internet. From optimizing public transportation to finding solutions for urban planning, air pollution or forecasting systems, these data prove continuously their efficacy.

In the public sector, the *advantages* of Big Data could be classified into three major groups: Big Data Analytics, improvements in effectiveness, and enhancements of efficiency (Munné 2016, 197). Concrete examples of these are: citizen segmentation, citizen personalization, smart cities applications, cybersecurity, data sharing, open government, and improvement of the quality of many public services (such as health, education, and social services). The *relevance* of Big Data technologies in the public sector is easily seen in their applications (Giest 2017); their further development requires improvements in data analysis, analysis of natural language, predictive analytics, modelling tools, and pattern discoveries. *Constraints* on Big Data may be summarized as: the lack of prompt political decisions needed to benefit from Big Data in the public sector; the lack of training for personnel in the necessary skills for the collection, interpretation, and archiving of Big Data; the absence of a standard set of solutions for this field; and the lack of specific resources (Munné 2016, 199). The *challenges* of Big Data applications are threefold: Big Data management issues such as collecting, retrieving, processing, and interpretation of results; ensuring data quality (a sensitive problem, related to not only the quality of the results but also to the financial and time investments); and ethical and privacy issues (privacy protection together with the encouraging of data sharing and the proper access to data) (Fredriksson et al. 2017, 48). Numerous governmental operations have proven the efficacy of using Big Data (Kim et al. 2014), but there are also situations in which they could potentially undermine public objectives and raise new threats (Janssen and van den Hoven 2015; Margetts and Sutcliffe 2013; Clarke 2016). The limited guidance in terms of ethical, legal, and policy frameworks has often made things more difficult.

To "demystify" the Big Data concept, a lot of research has been done using practical approaches, though the perspective of public managers is relatively disregarded in the literature (Guenduez et al. 2019). Using interviews with officials, Klievink et al. (2017, 268) found three main types of uncertainty: about what kind of Big Data uses is appropriate for their organizations, about their internal capacity for the proper use, and about their own organizational maturity with respect to the analysis of Big Data. Both overestimations and underestimations of how Big Data shapes the public sector are frequently found in the literature. As Schintler and Kulkarni (2014) noticed, we must get a correct picture of Big Data in public sector, including the good, the bad and the ugly. There is always a "dark side of Big Data", which includes the misuse of social media, inaccurate

algorithms, faulty modelling, and the biases of automated decision-making (Picciotto 2020), therefore the *evaluation* becomes essential.

3. Epistemological Challenges

The possibility of handling large quantities of information has led to qualitative changes that include *epistemological transformations*. Some of the most important questions are: *Do we think in the same way when we deal with Big Data? Do we form knowledge in the same way as we did before Big Data?* In 2008, Chris Anderson wrote a seminal article in *Wired* that raised several significant questions about the power of Big Data that can be summarized in just one: *will it bring about the end of theory*? Anderson analysed the ways in which large amounts of information are *firstly* mathematically treated while their context is established *later*. In the "petabytes age", the numbers seem to be enough to determine trends or patterns, sometimes without semantic analysis or causal judgment. In other words, what it is now considered merely "good enough" could eventually replace the classical model of scientific research based on hypotheses, tests, and models. In the information era, *correlation* seems to be a sufficient alternative to strong *causality*. The emphasis is on the way things are and not necessarily on the reasons behind. Every time something is gained, something else is lost: even if Big Data could offer macro-level patterns, they might not bring accuracy or insights on the micro-level. In philosophy, the concept of causality has raised fierce debates over time but represents a good manner of ratiocinating. In a certain manner, we could say that people were educated to search for cause-and-effect as an epistemological ground. By contrast, in a Big Data system, "we won't have to be fixated on causality; instead, we can discover patterns and correlations in the data that offer us novel and invaluable insights. The correlations may not tell us precisely *why* something is happening, but they alert us *that* it is happening" (Mayer-Schönberger and Cukier 2013, 26, authors' emphasis). General directions replace the in-depth examination of a topic, *what* substitutes *why,* trends supersede exactitude. In my view, for some analyses the mode of datafication (extracting general patterns and making predictions) could be more than sufficient, but for others the classical model of research must be applied (finding subtle explanations). Big Data possess the quality of granularity, and that allows a major level of clarity. The shifts in organizing research are the transition from small sets of data to massive quantities of information, with its corollary, the passing from sampling to the analysis of big data, as well as the recognition of the "messiness" of data and the crediting of correlation rather than causation (Mayer-Schönberger and Cukier 2013, 34-35). The intricate traits of Big Data modify some approaches but do not kill the theories. These methodological perspectives are founded on theories, and

the findings remain shaped by our choices (Mayer-Schönberger and Cukier 2013, 116; Boyd and Crawford 2012, 667). Nevertheless, "though it may seem counterintuitive at first, treating data as something imperfect lets us make superior forecasts, and thus understand our world better" (Mayer-Schönberger and Cukier 2013, 68).

Nevertheless, a "big data state of mind" is implied by using large datasets in decision-making processes such that this "underlying epistemology, rather than simply the bigness of datasets, affects the relationship between regulators and regulated entities, and the regulatory process at large" (Kempeneer 2021). Precisely from this reason, accountability and transparency are critical in using Big Data. As Kitchin (2014) stated in his seminal paper, the epistemic positioning is the main factor that differentiates Big Data from regular data and not their quantity. Big Data create a new framework through which we try to find the meaning of things or processes and "rather than testing a theory by analysing relevant data, new data analytics seek to gain insights 'born from the data'" (Kitchin 2014, 2). In this dynamic, the "dataism" contribute with a supplemental trust in the accuracy and objectivity of the information and algorithms, constructing an "algorithmic culture", with different ways of thinking and new practices.

4. Big Data: Whose Voice?

What are the consequences of this new model of thinking and analysing reality? For what citizens are the general directions extracted from a specific set of data representative? If some regulations are made based on the digital exhaust – the digital trail or fingerprints that a person creates because of his or her interaction with sites or online services – how could they be appropriate for individuals who do not use Internet? In this respect, a new form of digital divide emerges between "the Big Data rich and the Big Data poor" (Boyd, Crawford 2012: 674). Participation, access, and the interpretation of data are not always equally distributed, and these inequalities should be considered, especially if they produce biases. Digital divides have five dimensions: technical means, autonomy of use, use patterns, social support networks, and skills needed to effectively use online platforms (DiMaggio and Hargittai 2001). Inequalities do not appear only when some individuals or populations do not have Internet connectivity or smart devices. A strong discrepancy in using new media known as 'the second-level digital divide' is related to content creation and users' online abilities (Hargittai 2002). This situation relates to level of participation and has a great impact on the citizen online voice. The difference in online presence and skills will appear also at the level of representation when general directions are interpreted within Big Data. The importance of digital

literacy is obvious: technical access must be supported with effective education if citizens are to acquire specific digital competences. Many categories such as the homeless, elderly, poor, or ill people could be underrepresented in online data and their subsequent analyses. We may call it a form of the fallacy of hasty generalization when the conclusions derived from a set of information collected from a specific site are considered representative for all the population. As an example, even if in some countries X (formerly Twitter) is underused compared to other social networking sites, it has become very influential in indicating possible policy modifications. To extend findings from a specific online public to the general populace could lead to interpretative biases, neglect of some categories of citizen, and undemocratic measures. The issue of *representation* is central in public affairs and Big Data revitalizes the question of what citizen voices are really heard (Mergel et al. 2016, 935). The public policies should not favour people who have an online presence to the detriment of 'offline' silent citizens because it is possible that "while public administrators may know too much about some people, they may know too little about others and, thus, may potentially make wrong decisions about what and how public programs and corresponding services should be provided" (Guenduez et al. 2019, 2). It is also important to value *small data* for clarifying certain situations and obtaining precise answers (*data thickness*). In Geertz' tradition, the breadth of data should be complemented by their depth; thick data could resolve the context-loss of Big Data and bring out people's stories and emotions (Wang 2016). In a world dominated by massive amount of information, the relevant and successful ideas can rather come from "small patterns" (Floridi 2014), as we can see in branding and business (Lindstrom 2016).

5. Narrative versus database

Big Data brings into foreground *the dichotomy between narratives and databases*. Before the Big Data age, public affairs administrators created narratives for citizens based mainly on people's needs. They constructed causal explanations and models of how things should work. By contrast, databases are forms of structured data, and thus they allow information to be organized in categories according to different criteria. They are central to the computer age, a 'new symbolic form' based on algorithms and ready to be used for search or retrieval. A database works by parsing information, and problematic situations come from indeterminate data that do not fit in the predetermined categories or are borderline: should they be erased or made to have a null value? (Hayles 2012). Narratives and databases are generally competing cultural forms, or "natural enemies" (Manovich 2001, 225), but they could be seen as complementary. We need databases to tackle

massive amounts of information, but we also need narratives to understand complex relationships. The possibilities of Big Data are real – better targeting, enforcing participation, immediate insights in public' opinions, beliefs, behaviours – but inclusion must be ensured. A special emphasis must be also put on *evaluation* and *theory* in constructing the right framework for analysis. Without them, Big Data cannot reach its potential and, on the contrary, could generate many misunderstandings. Big Data, "being theory free, it cannot improve understanding of the world or infer causality. Being only effective for simple systems, consistent over time, it has limited predictive capacity in complex, changing, and volatile social environments" (Picciotto 2020,178). Proficiency in data analytics must be coupled with a refined sense of theory and evaluation, as well as ethical and legal commitments.

6. Conclusions

In the future, public affairs "will rely upon technology – digital and social media, real-time data, sophisticated algorithms, controlled vocabularies/living taxonomies, and emerging versions like artificial intelligence (AI) and natural language processing (NLP)" (Fleisher and McGrath 2020, 8), even if the current innovation of tools especially designed for public affairs has been "more limited than anticipated" (Fleisher and McGrath 2020, 6). At the same time, human involvement is not diminished in where digital technologies seem to occupy the very centre (Fleisher and McGrath 2020, 8). New media cannot replace human activities in the public sphere, but they could bring increased speed and efficiency. Public affairs are still a vocation in which practitioners must perform complex activities, with a growing level of interdisciplinary tasks. The necessity of theory, interpretive frameworks, and evaluation remain for the specialist. The technical competencies do not work alone; on the contrary, they must be supplemented with non-technical abilities. Big, open, and small data are needed to create the proper lens for understanding, interpreting, and ethically implementing strategies.

References

Anderson, C. 2008. "The end of theory: The data deluge makes the scientific method obsolete". *Wired* (23 June). Available at https://www.wired.com/2008/06/pb-theory/.

Australian Government, Department of Finance and Deregulation. 2013. "The Australian Public Service Big Data Strategy" (March). Available at https://www.mcguinnessinstitute.org/.

Boyd, D. and Crawford, K. 2012. "Critical questions for big data". *Information, Communication & Society* 15(5): 662-679. DOI: 10.1080/1369118X.2012.678878.

Chen, Y.-C. and Hsieh, T.-C. 2014. "Big data for digital government". *International Journal of Public Administration in the Digital Age* 1(1): 1-14. DOI: 10.4018/ijpada.201401010.

Clarke, R. 2016. "Big data, big risks". *Information Systems Journal* 26(1): 77-90. DOI: 10.1111/isj.12088.

Cunningham, S. W. and Thissen, W. A. H. 2014. "Three business and societal cases for big data: which of the three is true?". *IEEE Engineering Management Review* 42(3): 7-9. DOI: 10.1109/EMR.2014.2341476.

Desouza, K. C. and Jacob, B. 2017. "Big data in the public sector: Lessons for practitioners and scholars". *Administration & Society* 49(7): 1043-1064. DOI: 10.1177/0095399714555751.

DiMaggio, P. and Hargittai, E. 2001. "From the 'digital divide' to 'digital inequality': Studying Internet use as penetration increases". *Working Papers* 15. Princeton University, Woodrow Wilson School of Public and International Affairs, Center for Arts and Cultural Policy Studies. Available at https://digitalinclusion.typepad.com/.

Fleisher, C. and McGrath, C. 2020. "Public Affairs: A field's maturation from 2000+ to 2030". *Journal of Public Affairs* 20(3), 20:e2218. DOI: 10.1002/pa.2218.

Floridi, L. 2012. "Big Data and Their Epistemological Challenge". *Philos. Technol.* 25: 435-437. DOI: 10.1007/s13347-012-0093-4.

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., Gnanzou, D. 2015. "How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study". *International Journal of Production Economics* 165: 234-246. DOI: 10.1016/j.ijpe.2014.12.031.

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): 39-61.

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

Guenduez, A.A., Mettler, T., Schedler, K. 2020. "Technological frames in public administration: What do public managers think of big data?". *Government Information Quarterly* 37(1). Available at http://www.sciencedirect.com/.

Hargittai, E. 2002. "Second-level digital divide: Differences in people's online skills". *First Monday* 7(4). Available at https://doi.org/10.5210/fm.v7i4.942.

Hayles. K. N. 2012. *How We Think: Digital Media and Contemporary Technogenesis*. Chicago and London: The University of Chicago Press.

Hu, Q. 2018. "Preparing public managers for the digital era: incorporating information management, use, and technology into public affairs graduate curricula". *Public Management Review* 20(5): 766-787. DOI: 10.1080/14719037.2017.1327180.

Janssen, M., van den Hoven, J. 2015. "Big and open linked data (BOLD) in government: a challenge to transparency and privacy?". *Government Information Quarterly* 32(4): 363-368. DOI: 10.1016/j.giq.2015.11.007.

Janssen, M., van der Voort, H., Wahyudi, A. 2017. "Factors influencing big data decision-making quality". *Journal of Business Research* 70 (Supplement C): 338-345. DOI: 10.1016/j.jbusres.2016.08.007.

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. DOI: 10.1016/j.giq.2021.101578.

Kim, G. H., Trimi, S., Chung, J. H. 2014. "Big-data applications in the government sector". *Communications of the ACM* 57(3): 78-85. DOI: 10.1145/2500873.

Kitchin, R. 2014. "Big Data, new epistemologies and paradigm shifts". *Big Data & Society*, 1(1). DOI: 10.1177/2053951714528481.

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. DOI: 10.1007/s10796-016-9686-2.

Lindstrom, M. 2016. *Small data: the tiny clues that uncover huge trends*. New York: St. Martin's Press.

Liu, S. M. and Yuan, Q. 2015. "The Evolution of Information and Communication Technology in Public Administration". *Public Administration & Development* 35(2): 140-151. DOI: 10.1002/pad.1717.

Manovich, L. 2001. *The Language of New Media*. Cambridge, Massachusetts: The MIT Press.

Margetts, H. and Sutcliffe, D. 2013. "Addressing the policy challenges and opportunities of 'Big data'". *Policy & Internet* 5(2): 139-146. DOI: 10.1002/1944-2866.POI326.

Mayer-Schönberger, V. and Cukier, K. 2013. *Big data: A revolution that will transform how we live, work, and think*. Boston, New York: Houghton Mifflin Harcourt.

Mergel, I. 2016. "Big data in public affairs education". *Journal of Public Affairs Education* 22(2): 231-248. DOI: 10.1080/15236803.2016.12002243.

Mergel, I., Rethemeyer, R. K., Isett, K. 2016. "Big data in public affairs". *Public Administration Review* 76(6): 928-937. DOI: 10.1111/puar.12625.

Metcalf, J. and Crawford, K. 2016. "Where are human subjects in Big Data research? The emerging ethics divide". *Big Data & Society*, 3(1). DOI: 10.1177/2053951716650211.

Mullich, J. 2013. *Closing the big data gap in public sector*, *Survey Report*. New York: Bloomberg Businessweek Research Services.

Munné, R. 2016. "Big data in the public sector". In *New Horizons for a Data-Driven Economy. A Roadmap for Usage and Exploitation of Big Data in Europe*, edited by J. M. Cavanillas, E. Curry, W. Wahlster, 195-208. Cham: Springer. DOI 10.1007/978-3-319-21569-3_11.

Picciotto, R. 2020. "Evaluation and the Big Data Challenge". *American Journal of Evaluation* 41(2): 166-181. DOI: 10.1177/1098214019850334.

Schintler, L. A. and 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.

Snow Bailard, C. 2017. "Public Affairs, Digital Media, and Tech Trends". In *The Sage Handbook of International Corporate and Public Affairs*, edited by. P. Harris and C.S. Fleisher, 239258. Los Angeles: Sage.

Surowiecki, J. 2005. *The wisdom of crowds*. New York: Anchor Books.

The European Parliament and the Council of the European Union. 2019. "Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information" (26 June). *Official Journal of the European Union*. Available at https://eur-lex.europa.eu/.

The Government of Japan. 2013. "Declaration to be the World's Most Advanced IT Nation" (14 June). Available at http://japan.kantei.go.jp/.

The White House. 2012. "Obama Administration Unveils 'Big Data' Initiative" (29 March). Available at https://obamawhitehouse.archives.gov/.

The White House. 2014. "Big Data: Seizing Opportunities, Preserving Values" (May). Available at https://http://www.whitehouse.gov/.

Wang, T. 2016. *Why Big Data Needs Thick Data*. Available at https://https://medium.com/.

Yiu, C. 2012. *The big data opportunity: Making government faster, smarter and more personal*. London: Policy Exchange.