Ministerstwo Nauki i Szkolnictwa Wyższego
Doskonała Nauka
UTH UCZELNIA TECHNICZNO-HANDLOWA im. H. Chodkowskiej w Warszawie

1st International Scientific Conference “The multidimensionality of cybersecurity”

Future trends in big data security: a public governance perspective

Authors

Pietro Pavone[a]

Francesco Zappia[b]

Affiliation

[a] Department of Political Sciences
University of Naples Federico II
Naples, Italy

[b] Decision Lab
Department of Law, Economics and Human Sciences
University Mediterranea of Reggio Calabria
Reggio Calabria, Italy

Abstract

Big data is crucial for decision-making in contemporary societies, with clear connections at the level of public governance. It is even more true in an economic context characterized by social phenomena of growing complexity and interconnection, which develop according to the logic of reciprocity and interdependence. However, the collection and processing of enormous amounts of data can lead to beneficial but also harmful outcomes for the community. In cyberspace, risks and opportunities multiply. This study questions the possibility of distinguishing risks and obstacles that could alter the process of creating public value for the well-being of communities through cyber disvalue practices. For research and practitioners, this analysis could constitute a theoretical basis for relevant reflection to identify correct strategic data management strategies.

Keywords: bigdata; cybersecurity; public value; public disvalue.

Introduction

Public administrations provide complex outputs to satisfy individual and collective needs (Bozeman, 2007), aiming to create public value (Moore, 1995) through services, regulations, laws and other administrative actions.

Public value continues, however, to be a partly ambiguous and highly debated concept. In general terms, it can be understood as the set of citizens' expectations concerning the public policies and services they receive, thus raising the question of the social recognisability of the value generated by the administration. While the doctrinal reflection discusses new frameworks to improve the PA's ability to manage value through quantification and visualization models, the opportunities to create value starting from data are growing just as the pitfalls associated with the possibility of destroying value through fraudulent use are multiplying. some data (Rawat *et al*., 2019; Alani, 2021).

The ability of bodies and institutions to deal with the social problems of a community, developing new solutions and/or anticipating governance actions to prevent future critical issues, implies policies and programs that increasingly presuppose an evidence-based and less ideological, inclusive and participatory, with a long-term time horizon.

The strategic use of data - interpreted as an asset of public interest - represents a "great social promise" (Schintler and Kulkarni, 2014). The economic and social potential of data is, therefore, enormous, with new and additional challenges connected to the use of big data in cybersecurity (as a set of countermeasures, strategies, and standards that are used to prevent, detect, and defend against any vulnerabilities against system, organization network, or the Internet in the cyberspace).

This paper aims to analyze the paradigm in which the challenge of implementing data collection, processing and storage policies takes place with specific reference to the public sector. In particular, it involves addressing the issue of digital sovereignty as a need to affirm the role of governments, which must regain a position of centrality concerning the framework that loomed in the years preceding this challenge, in which the detention and use of the data were in the hands of a few large entities who achieved, and continue to achieve, a dominant position, economically advantageous for maintaining power even compared to the governments themselves.

The phenomenon of big data represents a fundamental axis to face the challenge of digitalization of the public sector which, for a few decades now, has been committed to revolutionizing strategies starting from the data archiving phase (from paper to cloud) in strict compliance with the legislation national reference, or the supranational one, but also by the strategies of other world powers.

Public organizations are, by definition, completely static systems; indeed, they should be able to change and expand their capabilities based on the necessary response and depending on the problems that the changing paradigm poses. Therefore, there has been a real substitution effect of private consultancy groups that have been involved in strategic political decisions and have even developed the ability to influence public policies. In this context, the scenario represented by the phenomenon of Big Data and, in particular, of Big Data Analytics will continue to propose a fascinating, as well as complex, challenge, as it focuses on the need to make profound changes, mainly because it will involve changing made man's relationship to information.

Methodological approach

Although presenting clear, practical implications, this contribution proposes a predominantly theoretical approach. In particular, an exploratory-descriptive research methodology is used, suitable for framing the topic of the use of big data for public security purposes according to a conceptual construct that takes into account the following perspectives: private condition of the data and public benefits (value) extractable from private data for defence and security purposes. The common thread of this study is represented by the effort to delve deeper into the proposed themes through a public management and governance perspective. The hypotheses on which the work is based and developed can be defined as follows:

The ultimate aim of the contribution is to create the necessary theoretical premises to answer the following research question (RQ):

RQ: Is it possible to trace possible future trends in the development of big data security based on the theoretical-practical evidence known so far?

The motivations for the research questions emerge from the following:

- the extreme relevance of the topic, both in the more developed and emerging economies and its new declinations in the digital age;
- non-exhaustiveness of existing interpretative models.

Big data concept

The complexity of big data is defined through the so-called 4Vs: 1) volume, 2) variety 3) speed 4) truthfulness. Some primary sources of big data are business transactions, computer networks, telecommunications networks, healthcare, finance, social media, bioinformatics, e-commerce, and surveillance. More recent studies have multiplied the Vs: Moro Visconti and Morea, 2019, for example , even talk about 10V of big data: volume, velocity, variety, veracity, variability, virality, viscosity, validity, visualization and value. However, for this work's purposes, we underline that every day, all individuals, firms and organizations produce big data.

The domain of Big Data Analytics (BDA) is not just about the sheer volume of data, but about extracting value from it: from data to decision. BDA is a powerful tool that allows us to make sense of the vast amount of information we have, turning it into actionable insights.

Big data in a public dimension

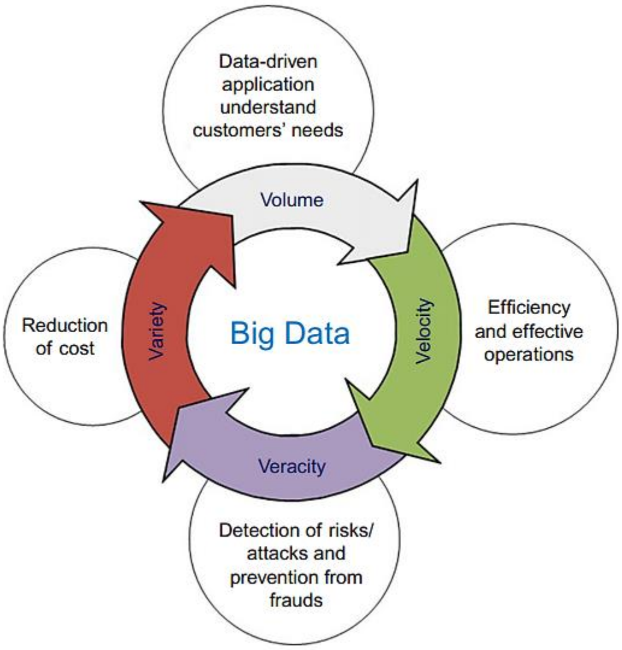
If we adopt a managerial conceptual angle,, big data is characterized above all by its granularity because it can detect specific aspects of social phenomena (George et al., 2014). Let us think about public policies and follow, for example, Pirog (2014). We can improve choices and public policies since we know the individual preferences of citizens better. For example, geospatial data (Wise & Craglia, 2010) can be crucial for public sustainability and social security policies. In political science studies, it is believed that you can learn more about the world by increasing your ability to collect and analyze data (Clark and Golder, 2015).

Big data in cybersecurity

McNeely and Hahm (2014) define the big data phenomenon as “a multidimensional concept embracing technology, decision making, and public policy”.

Big data can lead to different benefits (economic and social), through different possible applications (Figure 1).

Figure 1. Big Data and analytics applications


Data-driven application understand customers' needs
Volume
Reduction of cost
Variety
Big Data
Velocity
Efficiency and effective operations
Veracity
Detection of risks/ attacks and prevention from frauds

Source: Doku, R., & Rawat, D. B. (2019)

All applications must ensure mechanisms capable of preventing cyber-criminal attacks (Harris, 2014).

If we think about cities, especially the smart cities of the future, the intelligent use of technology must first visualize the benefits and risks of using big data since the governance of big data systems remains and will always remain a human prerogative, not a machine's. This is undoubtedly the first key point.

In creating "intelligent" public options, a large amount of data is usually generated and risks to public safety also lurk in the nodes of this process aimed at creating public value. Then, is it possible to create public disvalue while looking for the best way to generate public value? Unfortunately, the answer is “yes, this risk exists”. How, therefore, can we prevent cyber threats connected to the massive and dynamic use of data for public purposes from discouraging the ability to visualize and make data-driven strategic choices?

Benefits and risks of data collaboratives

The literature initially considered the technical dimension of the big data phenomenon. Subsequently, it arrived at conceptualizations that highlighted the social dimension and the main social challenges of big datasets (McAfee and Brynjolfsson, 2012).

In this "socialization" of the topic, we can simultaneously find benefits and risks.

Furthermore, this simultaneity is connected to the data's definition and economic nature: data are non-excludable and non-rival goods (for economists) because one use does not exclude others (even at the same time and even for fraudulent or harmful purposes).

To create innovative data combinations, we need to share data (van den Broek and van Veenstra, 2018), crossing organizational boundaries and connecting private and public sectors: in this sense, “data sharing is the practice of making data available for use by others ” (Michener, 2015, p. 34).

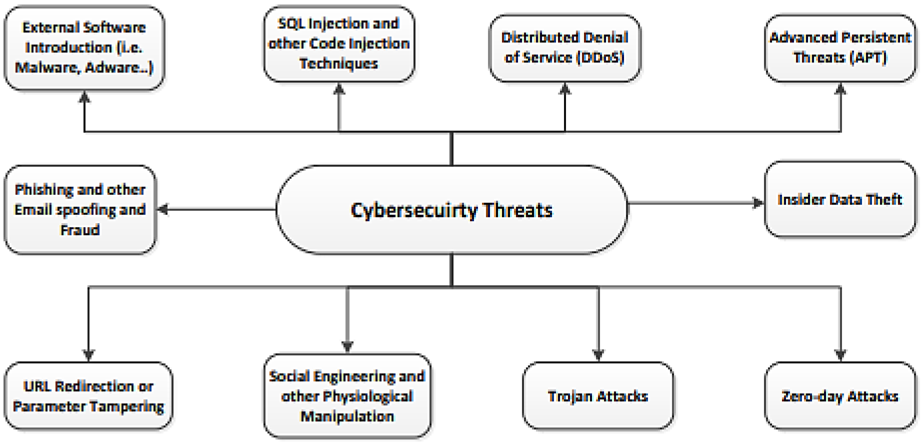
Some authors highlight that today's engagement in digital ecosystems is no longer a choice but a necessity for organizations (Prieelle et al., 2020). We can have different forms of data collaboration. However, the ultimate goal and meaning of these collaborative public governance schemes lies in accessing previously siloed data assets to leverage them in the public interest (Verhulst et al., 2019).

However, considering our crucial question (are there risks of public disvalue in public value initiatives?), what could be the pitfalls of these new data-driven governance trends?

Much research, for example, has focused on security in data sharing in the healthcare sector. The best solution to avoid crimes and harmful risks has been found in access control, data masking, and encryption techniques (Patil et al., 2017; Perez et al., 2017).

However, in more general terms, there is an accountability problem: “big data implies big accountability.” Algorithmic processes are enigmatic and obscure (De Laat, 2018). As algorithms become increasingly invisible and autonomous, it becomes harder for everyone to detect crime risks along the process chain. If we think about the classic cybersecurity threats known in the literature (Figure 2), we can construct the following Table 1, which offers an overview of the main barriers and risks that could be the fertile ground for criminal attacks in data collaborative contexts.

Figure 2. Cyber-security Threats


External Software Introduction (i.e. Malware, Adware..)
SQL Injection and other Code Injection Techniques
Distributed Denial of Service (DDoS)
Advanced Persistent Threats (APT)
Phishing and other Email spoofing and Fraud
Cybersecuirty Threats
Insider Data Theft
URL Redirection or Parameter Tampering
Social Engineering and other Physiological Manipulation
Trojan Attacks
Zero-day Attacks

Source: Nassar, A., & Kamal, M. (2021).

Table 1. Risks and barriers

- Account for the nature of the relationships among collaborators by a command-and-control hierarchy or by a principal-agent model (TECHNICAL) - Focus only on explicit standards of collaborative accountability (LEGAL) - Tension between different expectations (MOTIVATIONAL) - Uncertainties regarding the applicable liability regime (ETHICAL)
- Goal conflict can lead to conflicting performance indicators (MOTIVATIONAL) - Difficulty in attributing performance successes or failures to a single actor in the data ecosystem (TECHNICAL)

- New Data Analytics as a new form of NPM (MOTIVATIONAL) - Short-termism (data collected for short-term purposes) (TECHNICAL) - Risk of failing to prioritize data sharing over other pressing duties (MOTIVATIONAL) - Lack of resources (ECONOMIC)
- Risk that data providers do not only act as intermediaries in data sharing relationships (ETHICAL) - Risk of violating privacy laws (LEGAL) - Restrictive policies of data sharing (i.e., only open data model) due to a general sense of distrust or negative prior experiences (POLITICAL) - Risk of misleading data (errors in data collection, organization, and quality) (TECHNICAL)

Source: Authors' elaboration

Conclusion

Security in using big data is crucial to the future of data-driven public governance. Various actors (in the private and public spaces) must implement forms of collaborative governance in the context of data collaborations, first of all, building environments of transparency and accountability. Strengthening authentication schemes is urgently needed to improve data reliability.

Cyber threats and cybercrime risks multiply as the opportunities associated with the use of data increase. Future research should better study the practices of disvalue in the public and private sectors and - since the phenomenon is a transversal phenomenon - delve deeper into the risks that lurk in the intersection points between these two domains.

References

Alani, M. M. (2021). Big data in cybersecurity: a survey of applications and future trends. Journal of Reliable Intelligent Environments, 7(2), 85-114.

Bozeman, B. (2007). Public Values and Public Interest: Counterbalancing Economic Individualism, Georgetown University Press: Washington, DC, USA, pp. 1-214.

Clark, W. R., & Golder, M. (2015). Big data, causal inference, and formal theory: Contradictory trends in political science? PS: Political Science & Politics, 48(01), 65–70. https://doi.org/10.1017/S1049096514001759

De Laat, P.B. (2018), “Algorithmic decision-making based on machine learning from big data: can transparency restore accountability?”, Philosophy and Technology, Vol. 31 No. 4, pp. 525-541.

Doku, R., & Rawat, D. B. (2019). Big data in cybersecurity for smart city applications. In Smart cities cybersecurity and privacy (pp. 103-112). Elsevier.

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. Academy of Management Journal, 57(2), 321–326. https://doi.org/10.5465/amj.2014.4002

Harris, S. Securing big data in our future intelligent cities, in: Proc. of the IET Conference on Future Intelligent Cities, 2014, pp. 8–4.

McAfee, A. and Brynjolfsson, E. (2012). Big Data: The Management Revolution, Harvard Business Review, October, pp. 61-68.

McNeely, C. and Hahm, J. (2014). The big (data) bang: Policy, prospects, and challenges, Review of Policy Research, 31(4), pp. 304-310.

Michener, W.K. (2015). Ecological data sharing, Ecological Informatics, 29(1), pp. 33-44.

Moore, M.H. (1995). Creating Public Value, Harvard University Press, Cambridge, MA.

Nassar, A., & Kamal, M. (2021). Machine Learning and Big Data analytics for Cybersecurity Threat Detection: A Holistic review of techniques and case studies. Journal of Artificial Intelligence and Machine Learning in Management, 5(1), 51-63.

Patil, T.B., G.K. Patnaik, A.T. Bhole, Big Data privacy using fully homomorphic non-deterministic encryption, in: 2017 IEEE 7th International Advance Computing Conference (IACC), IEEE, 2017, pp. 138–143.

Perez, S., J.L. Hernández-Ramos, D. Pedone, D. Rotondi, L. Straniero, A.F. Skarmeta, A digital envelope approach using attribute-based encryption for secure data exchange in IoT scenarios, in: Global Internet of Things Summit (GIoTS), IEEE, 2017, pp. 1–6.

Pirog, M. A. (2014). Data will drive innovation in public policy and management research. Journal of Policy Analysis and Management, 33(2), 537– 543. http://www.jstor.org/stable/24033344. https://doi.org/10.1002/pam.21752

Prieelle, F. de, Reuver, M. de and Rezaei, J. (2020). The Role of Ecosystem Data Governance in Adoption of Data Platforms by Internet-of-Things Data Providers: Case of Dutch Horticulture Industry, IEEE Transactions on Engineering Management, pp. 1-11.

Rawat, D. B., Doku, R., & Garuba, M. (2019). Cybersecurity in big data era: From securing big data to data-driven security. IEEE Transactions on Services Computing, 14(6), 2055-2072.

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

van den Broek, T. and van Veenstra, A.F. (2018). Governance of big data collaborations: How to balance regulatory compliance and disruptive innovation, Technological Forecasting and Social Change, 129(April), pp. 330-338.

Verhulst, S., Young, A., Winowatan, M. and Zahuranec, A.J. (2019). Leveraging Private Data for Public Good: A Descriptive Analysis and Typology of Existing Practices, The GovLab, https://datacollaboratives.org/static/files/existing-practices-report.pdf (accessed 11 April 2024).

Wise, S., & Craglia, M. (Eds.). (2010). GIS and evidence-based policy making. CRC Press.