Aug 9th, 12:00 AM

The adoption of Big Data technologies - A challenge for National Statistics Offices

Fabio dos Santos Cardoso
University of Minho, id7039@alunos.uminho.pt

João Varajão
University of Minho, varajao@dsi.uminho.pt

Ana Carvalho
University of Minho, anac@eeg.uminho.pt

The adoption of Big Data Technologies - A Challenge for National Statistics Offices

Emergent Research Forum (ERF)

Fabio Cardoso
University of Minho
id7039@alunos.uminho.pt

João Varajão
University of Minho, ALGORITMI Center
varajao@dsi.uminho.pt

Ana Carvalho
University of Minho, NIPE
anac@eeg.uminho.pt

Abstract

National Statistics Offices (NSOs) integrate the global statistical network system. As it happens with other organizations, NSOs need to innovate in their technological structure to keep offering timely and high-quality official data. Big Data technologies are among the most relevant to improve the performance of NSOs. However, on the one hand, there is considerable variation among NSOs regarding the adoption of these technologies, which is a matter of concern. On the other hand, the phenomenon is not being addressed in the research literature. This study outlines research that aims to contribute to the understanding of how NSO organizations adopt and disseminate Big Data technologies in their main processes, including collecting, analyzing, and providing public statistics.

Keywords

Big Data technologies, Adoption, National Statistics Offices, NSO, Network.

Introduction

Official data is a major public asset to governments, societies, and citizens. Milan and Treré (2020) consider the official data gap as an essential aspect to governments dealing with contemporary challenges. However, data management can be a massive endeavor due to the high volume of data currently being generated, commonly referred to as Big Data. Urbinati et al. (2019, p.21) defines Big Data as *"(...) datasets that are large in volume, diverse in data sources and types, and created quickly, resulting in greater challenges to harvest, manage, and process them through traditional systems and capabilities*." Official data from health systems, economic performance, and poverty are examples of Big Data sets.

In this context, the National Statistics Offices (NSOs) are fundamental institutions since they provide official public statistics to countries and society. However, there are considerable inequalities among NSOs regarding the adoption of these technologies, which is a matter of concern. Furthermore, to the best of our knowledge, this phenomenon is not being addressed in the literature. The purpose of our research is to ascertain and explain how NSOs adopt and disseminate Big Data technologies in their main processes, which comprise collecting, analyzing, and informing official public statistics. This article presents the research proposal. The following section presents the background. Then, the research design is described. Finally, the expected contributions and future work are discussed.

Background

This research is grounded on two central bodies of literature: Big Data Adoption (BDA) and the Technology-Organizational-Environment (TOE) Framework. This section aims to help understand how the current literature supports the study of Big Data Adoption in NSOs. In this research, BDA represents an innovation for companies and their business, as proposed by Park and Kim (2019).

Big Data Adoption

The adoption phenomenon is a well-studied theme in Information System research (Brown and Venkatesh 2005; Trigo et al. 2013; Ho et al. 2020). Despite this fact, the combination of the search strings "Big Data" and "adoption" brings few results when searching in the Scopus service. Furthermore, when we search for BDA together with National Statistics Offices, there are no results.

Organizations integrating the network of NSOs work with mass processes in surveys, such as the price index and the census. If we look at other sectors that also deal with massive data, we find some studies. For instance, in the aviation sector, the case of the German company Lufthansa highlights that BDA success is strictly linked to CEO support (Chen et al. 2017). Other factors for successful BDA are, directly, IT infrastructure and, indirectly, the external environment (Chen et al. 2015). Regulation, compliance policies, and privacy also influence BDA process in different sectors (Gupta et al. 2018): energy and sustainability, lifestyle, healthcare, and disaster response. If those elements, regulation, compliance policies, and privacy, influence BDA in distinct application areas, it is plausible that they can also similarly affect BDA in NSOs.

In summary, the gap yet to be fulfilled in the literature refers to Big Data Adoption in National Statistics Offices. Those organizations are integrated into a global cooperation and mutual support network (The United Nations Statistical Commission 2020). This organizational design can carry specific features about BDA still not explored, such as the influence of a global system of partner organizations or the resistance from those partners to follow a worldwide guideline to adopt Big Data technologies.

The Technology-Organization-Environment Framework

There are several frameworks to study innovation and/or technology adoption. In this research, the Technology-Organization-Environment (TOE) model (De Pietro et al. 1990) was chosen. The TOE model presents a suitable design to explain innovation adoption at the organizational level (Baker 2012). Additionally, its graphic structure explains the linkages between internal and external adoption factors. These features make the TOE model appropriate to study Big Data Adoption in NSOs.

Several studies reinforced the use of the TOE model to explain BDA, such as Park and Kim (2019), and Gangwar (2018). According to Baker (2012), the TOE model analyses the firm under three perspectives: technological, organizational, and environmental. About technological innovation, the model proposes three different types: incremental, radical, and synthetic. All of them may result in the reinforcement or destruction of competencies, and they can impact the culture, processes, and the workers inside the organization, in a hard or soft way. Nowadays, BDA represents a technological innovation to NSOs, especially in the transition from the traditional survey techniques to new alternatives, such as the use of web scrapping in the consumer index price survey (Statistics Canada 2020).

In our research, the TOE model is considered relevant in its three dimensions: technological, organizational, and environmental. The first dimension refers to the features of Big Data as technological innovation. Internal resources and skills are potentially relevant variables during Big Data Adoption and implementation process in the organizational dimension. And the environmental dimension investigates the impacts of the organizational environment, including the network of National Statistics Offices, vis-à-vis Big Data Adoption.

Research Design

The research started with a systematic literature review (Savolainen et al. 2012), which established the study's boundaries. Using the set of strings ("Big Data," "Adoption," "Acceptance," "National Statistics Office," "NSO," "Instituto Nacional de Geografia e Estatística," "IBGE," "Office for National Statistic," "ONS," "Instituto Nacional de Estatística," "INE") in the Scopus and Web of Science databases, around 500 hundred references were identified. Scanning titles, abstracts, and keywords for relevance, reduced this number to around 100. The full reading of the papers supported this number of articles.

As a research theme, the *adoption of Big Data solutions by NSOs* presented coherence and applicability under the methodological criteria of Creswell and Creswell (2018). Based on this analysis, the central research question was defined as: Why and how do the National Statistics Offices adopt and disseminate Big Data technologies?

Since it is an unexplored theme, BDA in NSOs can be considered a new topic in Big Data studies. As such, the case study is an appropriate research design to guide this research process, considering the theme's originality (Eisenhardt, 1989). Another research design aspect to consider is the number of organizations to study. The multi-case study design (Yin 2014) seems to be adequate, and at least three NSOs will be studied: the Portuguese (INE), Brazilian (IBGE), and British (ONS) national statistics offices are the selected organizations to be analyzed. As nodes of the NSOs network, these organizations have similarities regarding the organizational mission and operations. They respond to their countries' respective governments as well as the United Nations Statistics Division - UNSD (The United Nations Statistical Commission 2020). The joint operation of NSOs within UNSD forms the Global Statistical System (GSS). In other words, the GSS natural configuration is a global network organization (Mizruchi and Galaskiewicz 1994). On top of that worldwide structure, some NSOs and UNSD play central roles in operational standardization and technology innovation processes. In contrast, others GSS members act under a follower perspective. The GSS represents a Core-Periphery network standard (Kurt and Kurt 2020). Its members keep working as a network (The United Nations Statistical Commission 2020) related to BDA.

As official data providers, NSOs direct their work processes, surveys, studies, and projections to guide and support public and private stakeholders' decision-making worldwide. NSOs can also be classified as *data actors* (Pappas et al. 2018*)*. Nowadays, those organizations deal with the challenge of losing relevance to societies in the face of new data providers (Milan and Treré 2020). Those new players are Tech Giants (TGs) such as Google, Apple, Facebook, and Alibaba. Some of them are under governmental pressure after provoking distortions in prices of specific markets, such as tourism (The Independent 2021) and digital advertisement (Financial Times 2021).

The data volume managed by the TGs through social media and the advancement of new data technologies pushed NSOs to deal with a dilemma: innovation or obsolescence (The United Nations Statistical Commission 2020). Some NSOs around the world have been trying to avoid obsolescence with BDA, as are the cases of CBS, the Netherlands' NSO (CBS 2017), and Statistics Canada (Statistics Canada 2020), with the adoption of web scrapping technologies in their surveys. Relevant types of Big Data technologies are cloud computing infrastructure, mobile data collection software, web scrapping, and scanner data applications, as defined by The United Nations Statistical Commission (2020). However, BDA is not a consensual option among the NSOs. Under 50% of NSOs have BDA projects deployed. Just 35% of them have a strategy for the use of BDA technologies.

The Brazilian, Portuguese, and British NSOs have different characteristics in terms of size, territorial coverage, technological development, and the legal constitution. Additionally, each organization has a distinct position in the GSS network. Firstly, the Brazilian Statistic Bureau, IBGE, provides information to approximately 212 million Brazilian citizens. It is a governmental organization subordinated to the Brazilian Federal Government (Instituto Brasileiro de Geografia e Estatísticas 2020). Inside the GSS network, this organization is in a peripherical position. The Instituto Nacional de Estatística (INE) reports its outputs and results to the Portuguese National Government. Its surveys cover around 10 million citizens (Instituto Nacional de Estatísticas 2020). This NSO is closer to the core of the GSS network. In this densely exchanging area of the GSS network, there are central nodes, such as UNSD, and more relevant NSOs, like ONS-UK. This NSO supplies data to 66 million people. It has a distinct feature: it is a nongovernmental organization with autonomy relative to the British Government (Office for National Statistics 2017).

The data sources for this qualitative research will include open-access documents and collected data from semi-structured interviews. Those interviews will be directed to decision-makers to complement or deepen the impressions obtained from the documental analysis. As part of the work plan, the prospective interviewees in NSOs will be professionals in five organizational positions: President (PR), Chief-Information-Office (CIO), Chief-Operation-Office (COO), Chief-Executive-Office (CEO), and Operational Manager (OM). Those organizational positions were defined based on the organizational structures of the three NSOs (Instituto Brasileiro de Geografia e Estatísticas 2020; Instituto Nacional de Estatísticas 2020; Office for National Statistics 2017). Considering previous studies with similar data collection instruments (Gangwar 2018; Park and Kim 2019), those five positions are considered the most significant to offer relevant data to answer the research question.

In order to develop a comprehensive understanding of how NSOs avoid or adopt and disseminate Big Data technologies, the TOE model will be used as a framework in the field research. Based on the information available in secondary data sources, it is proposed that the TOE elements can be translated to organizational

features related to NSOs. Table 1 exhibits some examples of the relation between the model's elements and the NSOs' elements.

TOE Elements	NSOs Elements
Technological innovation	Big Data technologies
Decision making	Acceptance or rejection; decision-making processes using Big Data technologies; centralized or decentralized decision-making
Organization	Internal organizational structure of NSOs
Formal and informal linking structures	Internal stakeholders, organization, formal and informal structures relevant for decision-making processes
Slack and Size	Budget, number of inhabitants, and covered geographic area
External task environment	Society, NSOs stakeholders, national economic conjuncture, Organizations' Network (GSS-UNSD and EUROSTAT)
Industry characteristics, market structure, and law	Official statistics, governmental data, regulation, compliance policies, and privacy
Technology support infrastructure	Internal, external or hybrid solutions; available skills and training in the labor market

Table 1. TOE Elements & NSOs features.

Conclusion

After finishing the literature review, this research will move on to the field study. This will include data gathering by collecting publicly available documentation and carrying out semi-structured interviews. The selected NSOs will also be visited. Direct observations and a second round of interviews will be performed during visits. Next, data analysis will take place. If needed, a second data gathering round will be carried out. The last phase will be the dissemination of results.

In terms of contributions, this research proposes to develop a comprehensive understanding of why and how Big Data technologies are adopted and disseminated in NSOs. It is expected that the results also contribute to extending the TOE framework application to the context of BDA adoption. Additionally, the literature on networks as organizational structures can also receive new insights. In this case, the proposed contribution can come from new findings on how nodes within the same network influence other nodes in the case of a technological adoption. Finally, the expected contribution for practice is to bring strategies and guidelines to support NSOs in their path for adopting Big Data technologies.

Acknowledgments

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scopes: UIDB/00319/2020.

REFERENCES

Baker, J. 2012. "The Technology–Organization–Environment Framework," in *Information Systems Theory*, Springer (ed.), New York, pp. 231–245.

Brown, S., and Venkatesh, V. 2005. "Model of Adoption of Technology in Households: A Baseline Model Test and Extension Incorporating Household Life Cycle," MIS Quarterly (29:3), pp. 399–426.

CBS. 2017. "Launching Center for Big Data Statistics," (https://www.cbs.nl/en-gb/our-services/innovation/nieuwsberichten/big-data/cbs-launching-center-for-big-data-statistics/), accessed November 15, 2020.

Chen, D., Preston, D., and Swink, M. 2015. "How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management," Journal of Management Information Systems (32:4), pp. 4–39.

Chen, H., Kazman, R., Schütz, R., and Matthes, F. 2017. "How Lufthansa Capitalized on Big Data for Business Model Renovation Big Data Deployment Gap," *MIS Quarterly Executive* (2017:March), pp. 443–448.

De Pietro, R., Wiarda, E., and Fleischer, Mitchell. 1990. "The Context for Change: Organization, Technology and Environment," in *The Processes of Technological Innovation*, L. G. Tornatzky and M. Fleischer (eds.), Lexington, MA: Lexington Books, pp. 151–175.

Creswell, J., and Creswell, J. D. 2018. *Research Design - Qualitative, Quantitative, and Mixed Methods Approaches*, (5th ed.), Thousand Oaks, CA: Sage Publications.

Eisenhardt, K. M. 1989. "Building Theories from Case Study Research," *The Academy of Management Review* (14:4), pp. 532–550.

Financial Times. 2021. "Facebook agrees to pay News Corp for content in Australia" (https://www.ft.com/content/3743eba8-dc91-47f8-add1-f2f9626f75dc) , accessed April 20, 2021.

Gangwar, H. 2018. "Understanding the Determinants of Big Data Adoption in India: An Analysis of the Manufacturing and Services Sectors," *Information Resources Management Journal* (31:4), pp. 1–22. (https://doi.org/10.4018/IRMJ.2018100101).

Gupta, S., Kar, A. K., Baabdullah, A., and Al-khowaiter, W. A. A. 2018. "Big Data with Cognitive Computing : A Review for the Future," *International Journal of Information Management* (42:July), pp. 78–89. (https://doi.org/10.1016/j.ijinfomgt.2018.06.005).

Ho, C. K. Y., Ke, W., Liu, H., and Chau, P. Y. K. 2020. "Separate versus Joint Evaluation: The Roles of Evaluation Mode and Construal Level in Technology Adoption," *MIS Quarterly: Management Information Systems* (44:2), pp. 725–746. (https://doi.org/10.25300/MISQ/2020/14246).

Independent. 2021. "Google fined $1 million for misleading French hotel rankings" (https://www.independent.co.uk/news/google-fined-1-million-for-misleading-french-hotel-rankings-google-french-paris-ap-hotels-b1802415.html), accessed April 20, 2021.

Instituto Brasileiro de Geografia e Estatísticas. 2020. "Structure". (https: //www.ibge.gov.br/en/institutional) November 15, 2020.

Instituto Nacional de Estatísticas. 2020. Organograma. (https://ine.pt/xportal/xmain?xpid=INE&. xpgid=ine_organograma_ine), accessed November 15, 2020.

Kurt, Y., and Kurt, M. 2020. "Social Network Analysis in International Business Research : An Assessment of the Current State of Play and Future Research Directions," *International Business Review* (29:September 2019), Elsevier. (https://doi.org/10.1016/j.ibusrev.2019.101633).

Milan, S., and Treré, E. 2020. "The Rise of the Data Poor: The COVID-19 Pandemic Seen From the Margins," *Social Media and Society* (6:3), pp. 1–5. (https://doi.org/10.1177/2056305120948233).

Mizruchi, M. S., and Galaskiewicz, J. 1994. "Networks of Interorganizational Relations," in *Advances in Social Network Analysis - Reserach in the Social and Behavioral Sciences*, S. Wasserman and J. Galaskievicz (eds.), London: Sage Publications, pp. 230–253.

Office for National Statistics. 2017. *ONS Organisational Charts.* (https://www.ons.gov.uk/), accessed November 15, 2020).

Pappas, I.O., Mikalef, P., Giannakos, M.N., Krogstie, J., & Lekakos, G. 2018. "Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies, " *Information Systems and e-Business Management*, (16:3), pp. 479–491.

Park, J. H., and Kim, Y. B. 2019. "Factors Activating Big Data Adoption by Korean Firms," *Journal of Computer Information Systems*, pp. 1–9. (https://doi.org/10.1080/08874417.2019.1631133).

Savolainen, P., Ahonen, J. J., and Richardson, I. 2012. "Software Development Project Success and Failure from the Supplier's Perspective: A Systematic Literature Review," *International Journal of Project Management* (30:4), pp. 458–469. (https://doi.org/10.1016/j.ijproman.2011.07.002).

Statistics Canada. 2020. "The Integration of Web-Scraped Data into the Clothing and Footwear Component of the Consumer Price Index", *Catalogue no. 62F0014M* (February), pp. 1-3

The United Nations Statistical Commission. 2020. "Background Document for the Note by the Secretary-General Transmitting the Report of the Global Working Group on Big Data for Official Statistics," (E/CN.3/2020/24). New York, United Nations.

Trigo, A., Varajão, J., Barroso, J., Soto-Acosta, P., Molina-Castillo, F. J., and Gonzalvez-Gallego, N. 2013. "Enterprise Information Systems Adoption in Iberian Large Companies: Motivations and Trends," in *Enterprise Resource Planning: Concepts, Methodologies, Tools, and Applications* (Vol. 2–3), pp. 1035–1052. (https://doi.org/10.4018/978-1-4666-4153-2.ch060).

Urbinati, A., Bogers, M., Chiesa, V., and Frattini, F. 2019. "Technovation Creating and Capturing Value from Big Data : A Multiple-Case Study Analysis of Provider Companies," *Technovation* (84–85:July), pp. 21–36. (https://doi.org/10.1016/j.technovation.2018.07.004).

Yin, R. K. 2014. *Case Study Research - Desing and Methods*, (5th ed.), London: Sage Publications.