Big Data Governance: *The Case of Mobile Positioning Data for Official Tourism Statistics In Indonesia*

Eko Rahmadian[1], Yulia Virantina[2], Daniel Feitosa[3], Andrej Zwitter[1], Titi Kanti Lestari[4]

[1] Campus Fryslan, University of Groningen, The Netherlands
[2] Statistics Indonesia, Indonesia
[3] Faculty of Science and Engineering, University of Groningen, The Netherlands
[4] Atmajaya University, Indonesia

Abstract:

The National Statistics Office (NSO) of Indonesia (Statistics Indonesia) has been using Mobile Positioning Data (MPD) to measure cross-border foreign visitors since 2016. Indonesia is one of the few countries that have already used MPD as one of its official statistics products, as it provides more accurate data with better coverage and timeliness on tourism arrival compared to the traditional method. Following its success, since 2018, research on the potential use of MPD has been expanded to other purposes, such as measuring domestic tourism and people's mobility in metropolitan areas. Not only the MPD but research on the potential use of other Big Data sources for official statistics has also increased response to the demand by related stakeholders and decision-makers. Following that, with the development and application of Big Data in various sectors and purposes, including official statistics, the role of Big Data governance is becoming increasingly important. Big Data governance is a holistic approach that allows the harmonization of people, methods, tools, and technologies to deal with structured and unstructured data. Big Data governance is also a new stage in the development of data governance, especially in exploring its theory and practice to improve organizational data management and utilization. Currently, despite the current success of the use of MPD, there are some challenges regarding Big Data governance that have possibly become threats to data sustainability and the entire data provision process. In this paper, we aim to investigate the issues and challenges of Big Data governance in the case study of MPD for tourism statistics in Indonesia. Our research aims to identify challenges in the dimensions of the big data governance framework, specifically in addressing issues on the role and communication among stakeholders, institutions or organizations, data quality, and regulatory compliance. To that aim, we conducted a field study in Statistics Indonesia, consisting of semi-structured interviews with related stakeholders. Through the result findings of our qualitative research on the MPD case study, we expect both to provide more insight and understanding of the urgency of big data governance and its framework for official statistics.

Keywords:

big data governance, mobile positioning data, official statistics, tourism statistics, data sustainability, sustainable tourism

1. Introduction

Since 2017, the potential use of Mobile Positioning Data (MPD) for tourism has been promoted by the World Tourism Organization (UNWTO) (Demunter, 2017). Meanwhile, Statistics Indonesia has been utilizing MPD to measure cross-border foreign visitors since 2016 (Noviyanti et al., 2020). MPD is a method of tracing coordinates from the cellular network, where the geography of

the cell is determined through the base radio station (Saluveer et al., 2020). Prior to the use of MPD, Statistics Indonesia relied on administrative data from the national immigration services to measure visitor arrivals in border areas through an annual survey, The adoption of MPD has provided more accurate data with better coverage and timeliness on tourist arrivals compared to traditional methods (Saluveer et al., 2020). Improving timeliness and coverage is crucial for statistics, as this MPD data is published as an official statistics product. Additionally, the use of MPD enables data disaggregation to be more precise at the regional level.

Following the success of its implementation, since 2018, the study on the potential use of MPD has been expanded to other purposes such as people mobility in metropolitan areas and domestic tourism. Furthermore, not only the MPD but the research on the use of other big data sources for official statistics has increased due to the demand by stakeholders and the need to provide better insight for decision-making at the national level. Big data refers to large, complex, structured, or unstructured data sets. One of the critical concerns in handling big data is the adoption of appropriate big data governance frameworks to (1) curate big data in a required manner to support quality data access for effective machine learning and (2) ensure the framework regulates the storage and processing of the data from providers and users in a trustworthy way within the related regulatory frameworks (both legally and ethically).

Currently, Statistics Indonesia is in the process of developing a big data governance framework to provide more trustworthy and reliable statistical products. Big data governance refers to managing huge volumes of an organization's data and exploiting it in its decision-making using different analytical tools (Al-Badi et al., 2018). Our paper aims to identify issues and challenges encountered by Statistics Indonesia regarding big data governance, specifically on addressing issues on the role of stakeholders, communication and coordination, and regulatory compliance. To achieve this goal, we conducted a field study in Statistics Indonesia, consisting of semi-structured interviews with related stakeholders, to identify the issues and challenges regarding big data governance on the MPD. Our paper is structured as follows: in the next section, we present the literature review on MPD and big data governance. We then proceed with the methodology and result findings. Finally, we present a conclusion and recommendations for future studies.

2. Literature review

2.1. The use of MPD for tourism studies

Historically, tourism statistics, including tourist volumes, expenditures, and tourist and trip-related characteristics, have been primarily collected through traditional surveys administered to a sample of households or individuals. These surveys have traditionally been considered the primary data source for domestic and outbound tourism statistics, as stipulated in the International Recommendations for Tourism Statistics 2008 (UNWTO, 2010) and Eurostat's Methodological Manual for Tourism Statistics (Eurostat, 2014). Additionally, border surveys are conducted to collect data on non-resident trips (Eurostat, 2014). However, surveys have several limitations, such as high cost, limited geographical scope, small sample size, and retrospective data acquisition, which can introduce bias to the results (Beaman, Huan &Beaman, 2004; De Cantis et al., 2015).

Recent advances in information and communication technology (ICT) have significantly changed tourism and travel behaviors. In response, researchers have turned to the increasing number of novel ICT-based data sources to study tourism. These sources can be used to analyze human activity or movement, making them useful for tourism research and statistics (Demunter, 2017). A review by Li et al. (2018) study on the application of big data to tourism research identified three primary sources of big data: tourists as users, devices or sensors used by tourists,

and various operations related to tourists (Li et al., 2018). Prior to this review, Demunter (2017) classified the most common sources of big data for tourism statistics as communication systems, the world wide web, business process-generated data, sensors, and crowdsourcing.

The use of location data from mobile phones in tourism studies has been subjected to active research since 2007 (Ahas et al., 2008). There are two means of obtaining mobile phone data, namely active and passive positioning. Active positioning refers to tracking mobile phones using specialized queries or using an installed application with the user's permission. However, this method is seldom used due to privacy concerns and the relatively small sample sizes it produces (Ahas et al., 2008). Alternatively, passive MPD has been commonly used in tourism research (Ahas et al., 2008). Passive MPD consists of location coordinates of mobile phones in a cellular network, automatically collected and stored by the mobile network operator (MNO) for customer billing, network maintenance, and performance monitoring purposes (Ahas et al., 2008). Call detail record (CDR), which includes active phone use data entries such as incoming and outgoing calls and sent messages (SMS, MMS), is the most common passive MPD set. Passive MPD can be used to measure the volume of tourists (arrivals and departures) and analyze tourist-and trip-related characteristics, such as country of origin, time, length, and spatial extent of visits, across all three forms of tourism (domestic, inbound, and outbound).

2.2. MPD for official statistics

The concept of "official statistics" encompasses various meanings, such as institution, process, and results (Rademarcher, 2019). In modern societies, official statistics are crucial in providing a fair and accurate representation of social, economic, and ecological phenomena through key figures, graphs, and other information (Turnpenny et al., 2015). The use of official statistics as an evidence-based policy is indicative of good governance characteristics.

Each country has the independence to organize its official statistics provision, known as National Statistical System (NSS). In most countries, statistical institutions are the producers of statistics. Using scientific, statistical methods, data is collected and processed in order to calculate condensed information which is made available to the general public. For instance, in Indonesia, based on Law number 16 of 1997, the task for official statistics provision is assigned to Statistics Indonesia, locally known as BPS. MPD has been one of the data sources for official tourism statistics since 2016. As an official statistics product, MPD should address the international standards, as well as international statistical principles and good practices implemented through the national statistical system using national legal frameworks and guidelines. For example, the UN Fundamental Principles of Official Statistics are: (1) Relevance, impartiality, and equal access; (2) Professional standards, scientific principles, and professional ethics; (3) Accountability and transparency; (4) Prevention of misuse; (5) Sources of official statistics; (6) Confidentiality; (7) Legislation; (8) National coordination; (9) Use of international standards; and (10) International cooperation.

2.3. Big data governance

The emergence of big data has presented new opportunities for the social, economic, and ecological domains, which have traditionally relied on official statistics provisions. However, realizing the full potential of big data requires addressing various challenges such as data processing, storage, access, and methodology. Collaborative efforts involving statistical producers, data holders, data users, and other stakeholders are necessary to achieve this objective (Struijs et al., 2014).

Related to this, big data governance assumes a critical role. It comprises processes, technologies, methods, and practices that facilitate the rapid discovery, collection, processing,

analysis, storage, and disposal of big data; while ensuring security, privacy, and cost efficiency (Malik, 2013). It also defines decision-making rights and responsibilities concerning the use of big data. Given that big data environments involve processing vast volumes of data within or outside an organization, effective data quality management, security, and ethical data processing are imperative. The integration of big data and business disruption, however, can heighten the frequency and magnitude of data breaches (Yang et al., 2019). Big data governance aligns people and technology towards averting costly data outages, promoting efficient management, and optimizing enterprise data utilization. Without adequate security controls, big data poses a significant security risk, with the scale and cost of the threat increasing with the volume of data collected. Therefore, the adoption of effective security measures to protect data at rest, on the network, and when leaving the network is mandatory. Moreover, access control during the analysis process and result provision must be implemented according to data sensitivity levels.

Establishing a robust big data governance framework is essential to enable effective data governance programs across the enterprise. This framework must possess the capacity to accommodate future growth while maintaining the flexibility to adapt to evolving data governance policies and implementations (Ghavarmi, 2020). Despite the crucial role that big data governance frameworks play in managing large data sets, few comprehensive frameworks have been proposed. This study refers to the work of Al-Badi (2018), who presented a big data governance framework that meets 87 percent of the ISO 8000 standard criteria. ISO 8000 is an internationally recognized standard designed to help organizations optimize their business process and control operational costs by improving data quality and governance. The framework of Al-Badi (2018) comprise eight components: (1)Communication and data management; (2) Identify organizations structure; (3) Stakeholders selection; (4) Data storage; (5) Big Data scope determination; (6) Measure and monitor quality; (7) Optimize and compute; and (8) Policies and standards setting.

3. Methods

This study employs a qualitative approach to address the research questions formulated above. In order to contextualize the governance of big data, including its application for official statistics, a theoretical framework was developed based on existing literature on big data governance dimensions (Al-Badi et al., 2018). To further explore the issues and challenges of MPD big data governance, semi-structured interviews were conducted with two groups of stakeholders comprising individuals with technological and non-technological backgrounds (n=13).

Semi-structured interviews were chosen as the primary method of data collection as they allow for an in-depth exploration of participants' perceptions, experiences, and opinions relating to large-scale data repositories and research (non) participation (Kvale, 2007). This is a verbal interchange where the interviewer attempts to get information from another person through a list of predetermined questions (Clifford, 2016). The interview topic guide covered relevant topics, and interviews were conducted until saturation was reached. The interviews were conducted by a trained member of the research team and lasted between 30-60 minutes. Audio recordings of the interviews were obtained with the participants' consent and subsequently transcribed and coded by the first author and checked by the second author. The complete study design protocol was assessed by the third and fourth authors and approved by the faculty's ethical committee.

4. Results

At present, Statistics Indonesia has no formal framework for big data governance. Although a draft has been prepared, it has not been finalized due to various reasons. Our research results could potentially contribute to developing this framework by addressing critical dimensions. In the

following sections, we present our findings related to the dimensions and challenges of MPD big data governance.

4.1 Policies and standard setting

Policies and standard settings are critical aspects of big data governance, and our study identified several issues related to laws and regulations in the context of MPD data collection. Personal data protection, big data access, and data interoperability among stakeholders are among significant challenges. The regulations of different government bodies often overlap, necessitating harmonization at the national and institutional levels involving key stakeholders, such as the Ministry of Planning, Ministry of Tourism, Ministry of Communication and Information, Ministry of State Enterprise, and Statistics Office.

Another issue related to law and regulation is the need to amend the law of statistics (Act 16/1997) to officially recognize big data as a source of official statistics in the country. Currently, big data is not considered a data source for official statistics based on law, which may limit its research, application, and development, not only for MPD but also for other big data sources. However, amending the law is a lengthy procedure involving political decisions both in the cabinet and in the parliament. Moreover, there is an ongoing debate about the use of big data for official statistics, with some arguing that the nature of big data is distinct from traditional statistics. To address this gap, Statistics Indonesia continues to engage with international organizations such as the UN Commission on Big Data and draw upon existing guidance at the international level.

4.2. Communication and data management

Ensuring the importance of MPD is a challenging task, requiring effective communication and political will at higher levels regarding data access, collection, and sustainability. Initially, MPD was used only for tourism statistics by Statistics Indonesia, but now is being employed for different purposes by various ministries, such as transportation, urban planning, and disaster management. Despite its wide benefits, certain key stakeholders are still resisting easing the data access of MPD. From the data producer side, it is expected to apply a government-to-government (G-to-G) mechanism for data provision. But currently, it follows a business-to-government (B-to-G) partnership model, which affects the process and cost for data access and other processes. Furthermore, data ownership is considerably low under this mechanism, which requires communication between the stakeholders to address this issue.

Another issue related to communication and data management is the importance of administrative and legal actors to ensure the success of data provision. It is required that the B-to-G mechanism has to implement the new regulation on e-procurement, including MPD. This new implementation has brought some problems with the contract, which lead to delays on certain phases, starting from the bidding process. This delay in the bidding process has a multiplier effect on the other phases, namely: data collection, processing, and analysis. To solve this issue, a new clause in the contract has been utilized that BPS could access the data, with or without the contract, for a certain period of time. This strategy is temporarily efficient and may not be a sustainable solution for tackling the issue in the long term.

4.3. Data storage

Other major issues that occurred are related to the challenges of data access, dependency, and data storage. These issues are not unique to Indonesia but also at the global level. Access to data is often restricted by MNO, which limits data analysis and exploration. In addition, data storage costs can also be an obstacle, particularly when dealing with a large volume of data. For

instance, the Statistics Office may require significant expenses for data storage depending on the length of the data retention period.

Furthermore, a threat occurred on data sustainability regarding uncertainty surrounding the availability of certain variables. MNO often restrict data access or provide limited access to certain variables due to data confidentiality, which can hinder research efforts. To address this challenge, there is a need for higher levels of regulation or laws that facilitate data access while preserving data confidentiality. Without such regulations or laws, these uncertainties could undermine data sustainability, particularly in the context of big data initiatives. Moreover, even after data access is granted, further regulation is needed to ensure data sharing, interoperability, and data ownership.

4.4. Measure and monitor quality

There are pros and cons on the use of big data for tourism or other official statistics, particularly when compared to traditional statistical methods. One argument against big data is that its unstructured nature sets it apart from traditional statistics. In addition, administrative challenges associated with big data provision have led to questions on the benefits of timeliness and coverage compared to traditional surveys. However, the use of MPD has emerged as a promising method for providing reliable tourism statistics. Research in years and implementation have demonstrated MPD's accuracy in predicting tourist movements at regional and provincial levels. This maturity is important as a consideration in assessing the value or importance of big data for official statistics. Nevertheless, there are limitations regarding its accuracy in predicting at the lower level, including errors in signal reception and the quality of addressing point of interest, which vary depending on the country and type of data. As such, MPD may be best suited to generate statistics at a higher level, such as municipalities. Overall, using big data in official statistics requires careful consideration of the benefits and challenges, as well as an understanding of its integration with traditional statistical methods.

4.5. Identify organizations structure

There are two opinions in the internal of the Statistics Office on the debate of optimal organization scheme for big data governance. The first proposes the formation of a dedicated unit for big data analysis, which would require internal restructuring. While the second suggests continuing with the current approach of forming task force teams based on projects. Furthermore, to improve the leverage on the potential of big data, there is a need to expand and advance data processing and analysis, which requires additional human resources. Also, time constraints have posed a challenge for data exploration and analysis, given factors such as the administrative process and heavy workloads of the people. Therefore, to address these challenges, it is important not only to consider the expertise of personnel but also their passion and dedication. Another potential solution is to engage with academic researchers or data scientists from the Institute of Statistics to contribute to expanded and advanced data processing and analysis.

4.6. Optimize and compute

MPD has been used as a data source for official tourism statistics in Indonesia and to aid decision-making on sustainable tourism and SDGs indicators through the use of certain MPD variables, such as signal type. During the early stages of the MPD initiative, Statistics Indonesia faced challenges related to infrastructure or ecosystem. As a solution, Statistics Indonesia partnered with MNO to leverage their storage spaces, servers, and command centers. This partnership allowed for successful implementation despite infrastructure and ecosystem challenges.

In addition, methodological challenges were also encountered during the beginning period on the use of MPD. That challenge involved determining basic tourism concepts by distinguishing between the usual and unusual environments of tourists. To analyze that concept and to be able to publish monthly official statistics, it needs at least a data set of one year backward, which consists of a huge volume of data for the analysis. Data access was also a challenge as Statistics Indonesia initially requested signaling data instead of CDR data which are only recorded when there is an activity such as a call, or SMS. MNO considered signaling data as garbage data, which would be deleted to save storage capacity.

Despite the challenges in technology and human resources, it seems Statistics Indonesia has the capacity. However, it is still less optimal. The organization lacks time to utilize the data for other analyses, such as points of interest, movement within the municipality, mode of transportation,and poverty analysis. For instance, based on phone credit data, it is possible to determine the economic level of people, or based on movement, whether they stay in one place for a prolonged period.

4.7. Big data scope determination

Determining big data scope or priority is important to ensure that the organization works efficiently. Given the scale of big data research and statistical products, it is impractical to prioritize everything. Thus, strategic planning is necessary to achieve the best results. Currently, big data initiatives in Statistics Indonesia need to be centralized. The absence of a roadmap causes uncertainty when identifying priorities. To this end, recent criteria have been considered for prioritizing big data initiatives, such as assisting basic statistics, fulfillment of direct presidential or governmental instructions, and national priorities. These criteria should also align with the One Data Indonesia program and the National Standard of Statistical Data (SDSN). In addition, prioritizing big data initiatives based on data sustainability. Ensuring data access and quality enable Statistics Indonesia to focus on big data research development. Despite these criteria, Statistics Indonesia permits all big data initiatives to thrive. Nevertheless, developing a big data roadmap is recommended to improve data governance and quality.

5. Conclusion

This study's objective is to comprehensively understand the challenges of big data governance in multiple dimensions, using the case study of MPD in Indonesia. Big data governance ensures data quality and promotes trustworthiness, consistency, and reliability in managing risks or challenges. Based on our findings, we propose to use a big data governance framework for official statistics to guide further efforts and applicability.

Our findings show that the main challenges facing MPD big data governance are the lack of law and regulations, issues with data access and cost, administrative process, and insufficient data exploration. These issues are interconnected, and other relevant topics, including communication, partnership, organization, human resources, and methodology, are also covered in our study. Successful big data governance requires a comprehensive understanding of all aspects, enabling organizations to identify opportunities, challenges, and requirements concurrently. To this end, we propose the use of a big data governance framework for official statistics based on the MPD case study, which is complex and applicable to other big data sources. In conclusion, ensuring data sustainability is a complex and multifaceted challenge that requires careful consideration of various legal, ethical, and technical issues.

References

1. Ahas R, Aasa A, Roose A, et al. (2008) Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management* 29(3): 469–486. DOI: 10.1016/j.tourman.2007.05.014.
2. Al-Badi A, Tarhini A and Khan AI (2018) Exploring Big Data Governance Frameworks. *Procedia Computer Science* 141: 271–277. DOI: 10.1016/j.procs.2018.10.181.
3. Beaman JG, Huan T-C and Beaman JP (2004) Tourism Surveys: Sample Size, Accuracy, Reliability, and Acceptable Error. *Journal of Travel Research* 43(1): 67–74. DOI: 10.1177/0047287504265514.
4. De Cantis S, Parroco AM, Ferrante M, et al. (2015) Unobserved tourism. *Annals of Tourism Research* 50: 1–18. DOI: 10.1016/j.annals.2014.10.002.
5. Demunter C (2017) *Tourism Statistics: Early Adopters of Big Data? : 2017 Edition*. 2017 edition. Luxembourg: Publications Office of the European Union.
6. European Commission. Eurostat. (2014) *Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics :Consolidated Report.* LU: Publications Office. Available at: https://data.europa.eu/doi/10.2785/55051 (accessed 21 November 2022).
7. Ghavami P (2020) *Big Data Management: Data Governance Principles for Big Data Analytics*. 1st ed. Boston: DE GRUYTER.
8. Khatri V and Brown CV (2010) Designing data governance. *Communications of the ACM* 53(1): 148–152. DOI: 10.1145/1629175.1629210.
9. Kim HY and Cho J-S (2017) Data Governance Framework for Big Data Implementation with a Case of Korea. In: *2017 IEEE International Congress on Big Data (BigData Congress)*, Honolulu, HI, USA, June 2017, pp. 384–391. IEEE. DOI: 10.1109/BigDataCongress.2017.56.
10. Lestari TK, Esko S, Saluveer E, et al. (n.d.) Indonesia's Experience of using Signaling Mobile Positioning Data for Official Tourism Statistics.: 13.
11. Li J, Xu L, Tang L, et al. (2018) Big data in tourism research: A literature review. *Tourism Management* 68: 301–323. DOI: 10.1016/j.tourman.2018.03.009.
12. Noviyanti I, Prabawa PD, Sari DP, et al. (2020) Towards big data as official statistics: Case study of the use of mobile positioning data to delineate metropolitan areas in Indonesia. *Statistical Journal of the IAOS* 36(4): 943–954. DOI: 10.3233/SJI-200750.
13. Radermacher, J. (2020). Official Statistics 4.0. Verified Facts for People in the 21st Century. Springer International Publishing. DOI: 10.1007/978-3-030-31492-7
14. Raun J, Ahas R and Tiru M (2016) Measuring tourism destinations using mobile tracking data. *Tourism Management* 57: 202–212. DOI: 10.1016/j.tourman.2016.06.006.
15. Saluveer E, Raun J, Tiru M, et al. (2020) Methodological framework for producing national tourism statistics from mobile positioning data. *Annals of Tourism Research* 81: 102895. DOI: 10.1016/j.annals.2020.102895.
16. Shoval N and Ahas R (2016) The use of tracking technologies in tourism research: the first decade. *Tourism Geographies* 18(5): 587–606. DOI: 10.1080/14616688.2016.1214977.
17. Tallon PP (2013) Corporate Governance of Big Data: Perspectives on Value, Risk, and Cost. *Computer* 46(6): 32–38. DOI: 10.1109/MC.2013.155.
18. Turnpenny, John R., Andrew J. Jordan, David Benson, and Tim Rayner. (2015). The Tools of Policy Formulation: An Introduction. In The Tools of Policy Formulation, ed. Andrew J. Jordan and John R. Turnpenny. Cheltenham: Edward Elgar
19. United Nations World Tourism Organization. (2010). International recommendations for tourism statistics 2008. New York. United Nations.
20. Yang L, Li J, Elisa N, et al. (2019) Towards Big data Governance in Cybersecurity. *Data-Enabled Discovery and Applications* 3(1): 10. DOI: 10.1007/s41688-019-0034-9.