Key Research Challenges of Incorporating Official Statistics with Big Data Analytics and Lean Six Sigma

Farah Al Jaghoub
Industrial Engineering and Engineering Management, University of Sharjah
Sharjah, UAE
u20103917@sharjah.ac.ae

Zehra Canan Araci
Assistant Professor
Industrial Engineering and Engineering Management, University of Sharjah
Sharjah, UAE
zaraci@sharjah.ac.ae

Concetta Semeraro
Assistant Professor
Industrial Engineering and Engineering Management, University of Sharjah
Sharjah, UAE
csemeraro@sharjah.ac.ae

Abstract

The amount of created and captured data has been increasing rapidly. Since 2018, the amount of data generated and collected surpasses 90% of the data produced in the previous years. To meet the 2030 United Nations Economic Commission for Europe (UNECE) statistical agenda, National Statistics Organizations (NSOs) must meet the required amount of data demanded through the modernization of statistics. Relevant studies showed the implementation of official statistics with lean six sigma, resulting in a 5% operational budget reduction. Moreover, integrating official statistics with big data analytics produced more efficient and timelier statistics. However, in this industry 4.0 era, there is a lack of practices in the integration of big data tools and technique with the official statistics. The reason might be the uncertainty on the challenges and potential solutions. Therefore, this paper systematically reviews and analyses the critical challenges and potential solutions for the integration of big data analytics and lean six sigma into the processes of generating official statistics. A comprehensive and systematic literature review was conducted. Findings showed that the challenges of integrating big data are lack of proper infrastructure, data quality and data privacy. On the other hand, challenges of integrating lean six sigma are lack of management support, lack of common goal and the hierarchical organization structure. This paper will contribute to future work by using the identified challenges of integrating official statistics with big data analytics and lean six sigma to develop and propose a framework which overcomes the critical challenges and maximize the benefits.

Keywords
Official statistics, GSBPM, Big data and Lean six sigma.

1. Introduction

Globalization and digitalization enabled access to massive data sets and eased the connectivity of advanced technologies. This led to supporting the digital transformation of conventional statistical processes using appropriate advanced technologies. Using big data analytics can allow originating dynamic data sources to improve and add value to National Statistics Organizations' processes (Kitchin and McArdle 2016).

National Statistics Organizations (NSOs) responsible for publishing official statistics which deliver critical quantitative and qualitative information that affect citizens' lives. This information is critical to fulfilling the needs of decision-makers, researchers, business users, policymakers, educators, and other interested parties (Radermacher 2020). NSO worldwide are the most trustworthy information and high-quality data providers, essential for the economy and society. Due to the advancement of technologies, reduced budgets, and statistical demand growth, NSOs face changes to do more achievements with few resources. Lean six sigma helps organizations obtain a competitive advantage by implementing improvement methodology to create new organizational processes and structures and even to manage the changes in an organization. However, there are some challenges in integrating lean six sigma methodology with official statistics that should be addressed (McSweeney and Moore 2015).

On the other hand, the technology is advancing very fast. The amount of data is collected is considerably larger than ever before. Handling this vast amount of data is quite challenging which also affects generating official statistics. Therefore, it is inevitable to integrate big data analytics into the official statistics processes. It will help reducing the processing time and cost while ensuring the superior quality in data output. To be able to provide a framework that integrates big data tools and lean six sigma approach into the process of generating official statistics, the challenges and the main causes of the problems should be identified. However, there is lack of research providing enough information about a list of challenges, the reasons, and potential solutions. Therefore, this paper aims to address this gap by answering the question:

RQ1) What are the key challenges of implementing big data analytics and lean six sigma to improve the processes of generating official statistics?
Following section provides a detailed information about the methodology of a systematic literature review. Section 3 presents the findings of the extensive literature review of incorporation between official statistics, big data analytics and lean six sigma. Section 4 summarizes the critical challenges under two categories: (1) big data analytics and (2) lean six sigma. Finally, in section 5, conclusions are formulated, and recommendations are provided for future work.

2. Methodology

This section will briefly present the process of conducting the systematic literature review. This paper selection on incorporating official statistics with big data analytics is based on the following query [("official statistics") AND ("big data analytics")], while the search query [("official statistics") AND ("lean six sigma")] is for selecting documents on incorporating official statistics with lean six sigma. Figure 1 illustrates the stages used for conducting the search with the number of papers found. The search found 164 documents, with 161 papers about official statistics and big data analytics. However, only three papers were found regarding integrating official statistics with lean six sigma. In addition, authoritative websites were used to capture the implicit information not academically published. Table 1 presents the official websites used to collect data for this work.

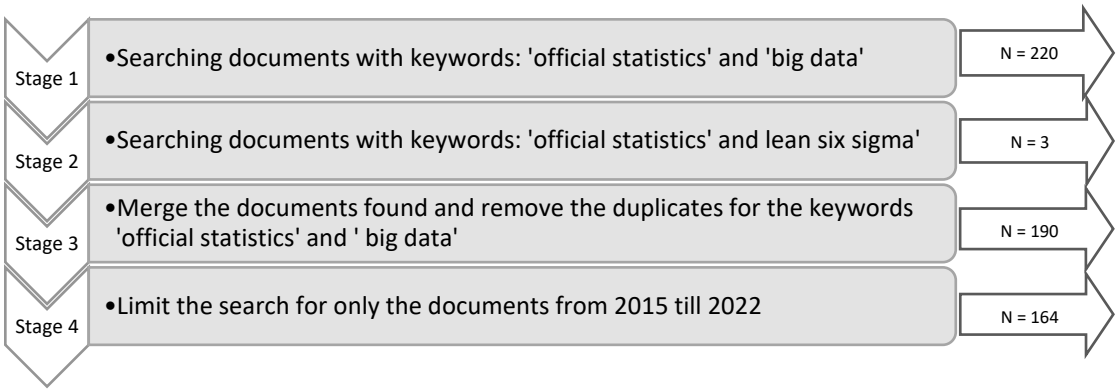

Figure 1. Stages of searching with number of papers found

Table 1. Authoritative websites used for the data collection

Authoritative Website	**Captured information**
The European statistical system website: https://ec.europa.eu/eurostat/cros/content/european-statistical-system-theme_en	- An overview of the implementation of GSBPM in the European statistical system
UNECE official website: https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1	- An understanding of GSBPM - Several case studies for countries that implemented GSBPM

VOSviewer text mining software was used in this work. VOSviewer is a scientific visualization tool utilized to analyze the bibliography data for the occurrence of the keywords of scientific publications by creating a map based on the network data (Van Eck and Waltman 2017). Figure 2 shows a screenshot of the results of the Scopus databases search of incorporating official statistics with big data analytics and lean six sigma. The circles indicate the keywords used by the scientific publications; the larger the circle is, the more scientific publications use it. Based on the VOSviewer map, official statistics have various big data applications, such as surveys, public policies, data mining, data privacy, and decision making. However, official statistics do not have many applications with lean six sigma due to a lack of documents in this regard. Also, VOSviewer could not be implemented for incorporating official statistics with lean six sigma because of the insufficient number of documents about the incorporation of official statistics and lean six sigma. Another analysis was conducted to identify the number of documents issued per year. Figure 3 shows that most papers were published from 2015 to 2021, with a peak value in 2021 with 46 documents.

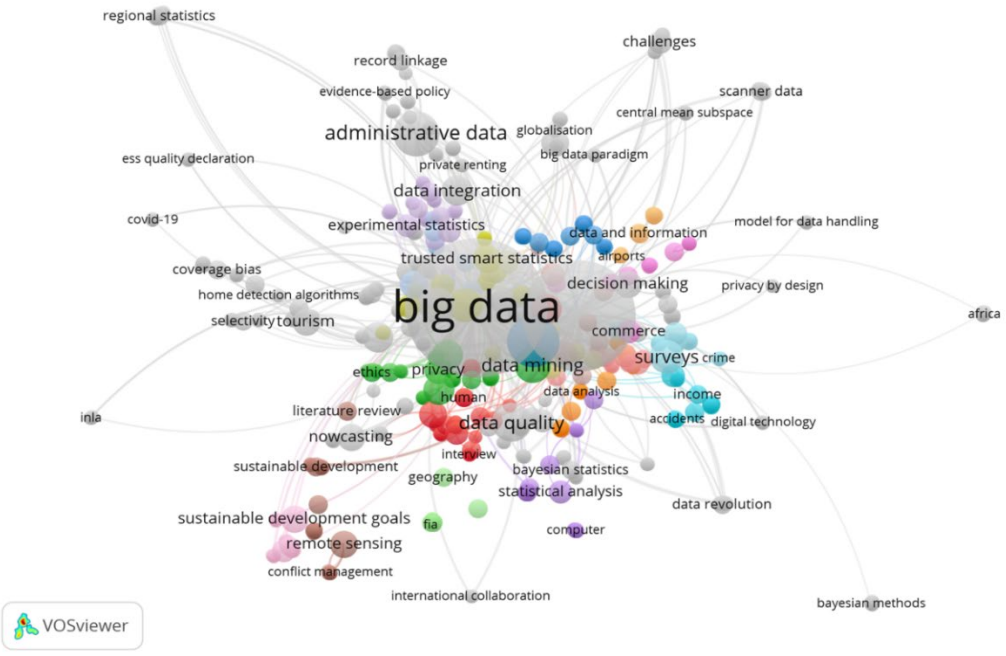

Figure 2. Results of Scopus databases search of Incorporating Official statistics with big data analytics and lean six sigma

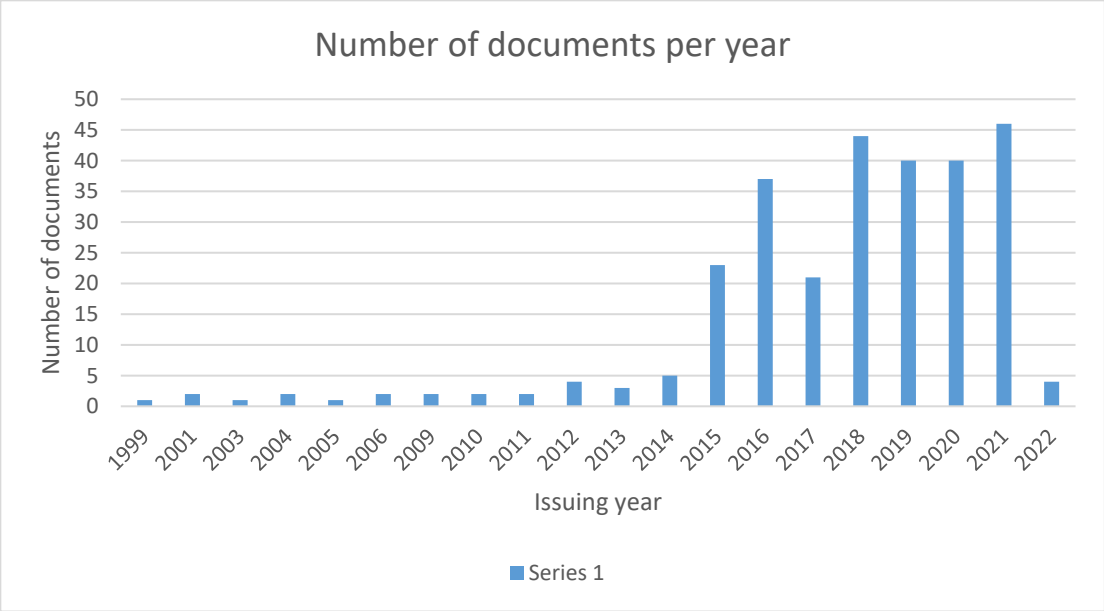

Figure 3. Number of documents per year

The systematic literature review was conducted to capture the concepts of official statistics, lean six sigma and big data analytics. Moreover, this analysis highlights the role of lean six sigma in official statistics and the role of big data analytics in official statistics to outline the associated key challenges within the defined roles.

3. Incorporating Official statistics with Big Data Analytics and Lean Six Sigma

3.1 Official Statistics Overview

Official statistics are statistics published by National Statistics Organizations (NSOs) that provide essential quantitative and qualitative information for decision-makers across major society aspects that affect citizens' lives. This information is crucial to fulfilling the needs of decision-makers, policymakers, researchers, business users, educators, and many other interested parties. The National Statistics Organizations publish most of the official statistics (Radermacher 2020). Radermacher (2020) defined in his book "Official Statistics 4.0" those official statistics are the output of a systematic process representing societal needs while considering the process's limitations. The production of official statistics serves its intended purpose as "products" to achieve the determined data quality and utility level. Additionally, official statistics can set up an infrastructure for the public that provides free-of-charge services for society. Disseminating official statistics is integral the process; therefore, formal cooperation and communication with media and journalists should be forged.

3.2 Big Data Overview

Globalization and digitalization enabled the world to extensively use telecommunication, the Internet, satellite, and many other technologies, increasing data collection over time. Based on Wadhera et al. (2021) study, 90% of all available data has been produced in the last four years. Data is massively amplified over time; thus, it needs the proper tools and techniques to deal with it in a time- and cost-efficient way (Goundar et al. 2021). The big data terminology was first introduced in 1997 by David Ellsworth and Michael Cox through NASA's Ames Research Centre (Cox and Ellsworth 1997). Big data is an extremely large amount of data obtained from conventional and digital sources. Data processing and analysis require advanced tools to acquire meaningful information from the data. There are many advantages to using big data tools and techniques, such as enhancements in productivity, improvements in the client journey, better effective utilization of products and services, and improvements in the decision-making process (Alwan and Ku-Mahamud 2020). Big data has three main characteristics; volume, velocity and variety which can be dubbed as the 3Vs. The rapid advancements in new technologies, system capabilities, and data storage capacities urged the need to change the definition of big data (Raja et al. 2020). Recent studies discussed the expansion of the 3Vs to 5Vs:

volume, velocity, variety, value, and veracity (Nti et al. 2022). The data's level of accuracy can explain value and veracity is the level of trust in the data (Chawda and Thakur 2016). Big data must be gathered, stored, processed, analyzed, and archived by specialized tools and techniques through a data life cycle. The first step in the data cycle is to acquire data from various sources such as the web, sensors, social media, mobile data, and the Internet of Things (IoT). The subsequent step involves storing the data in NoSql DBMS, RDBMS, and NewSql DBMS systems. Later, tools and techniques must process the stored data, such as Map Reduce, Hive, and Mahout. Followed by that, the data analysis is carried out by elastic Map Reduce and Big Query. Lastly, the data is archived in Amazon glacier and cloud backup. Figure 4 demonstrates the five steps incorporated within the big data cycle (Gupta et al. 2018).

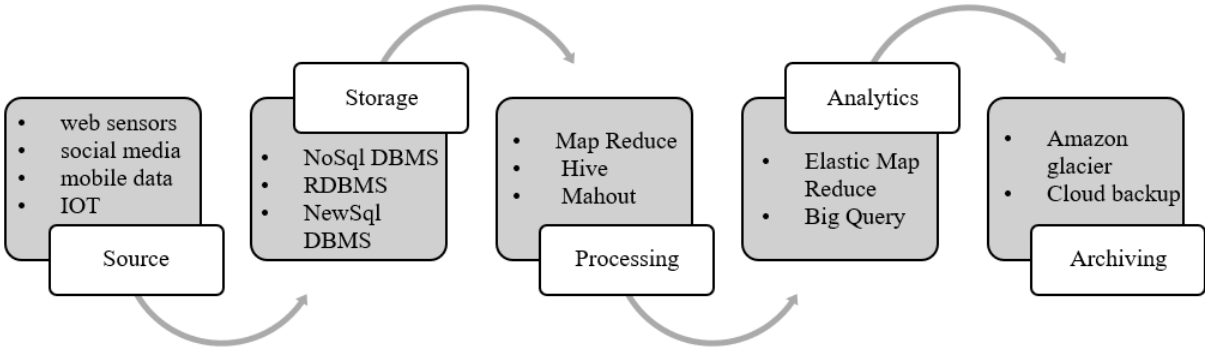

Figure 4. Big data life cycle

Big data tools and techniques can be classified into data gathering, monitoring, managing, and programming. Microsoft, Oracle, and IBM offer many big data open-source platforms, such as Spark, Storm, and MongoDB (Frizzo-Barker et al. 2016a). The most common crucial big data open-source platform analytic tools are MapReduce, a programming tool for huge parallel datasets, and the Hadoop, a MapReduce-based tool that deals with enormous datasets for data processing and monitoring (Adolph 2014). Many other open-source platforms for big data analytics, such as: Spark, Storm, and MongoDB (Frizzo-Barker et al. 2016b). Moreover, one of the tools and techniques used for big data analytics is data mining (Russom 2011). Machine learning an algorithm that can learn better and provide more results by capturing more data. Hence, machine learning is essential to data analytics (Zhou et al. 2017). The massive amount of big data is challenging to store, process, and analyze using conventional software and techniques. Employing machine learning algorithms with a significant volume of data effectively achieves prediction, classification, clustering, and recommendation (Wang and Alexander 2016).

Deep Learning is one part of machine learning consisting of complex layers to perform artificial neural networks, and machine learning is one section of artificial intelligence (Cully et al. 2015).Machine learning consists of several learning algorithms: supervised learning, unsupervised learning, artificial neural networks, and reinforcement learning. Figure 5 shows the differences between machine learning algorithms (Wang and Alexander 2016).

Supervised Learning	Unsupervised Learning	Reinforcement Learning
-Labeled type of data	-Unlabeled type of data	-No predefined data
-External training supervision	-No external training supervision	-No external training supervision
-Deals with regression and classification problem	-Deals with association and clustering	-Deals with reward-based problem
-Matching the input with the output	-Interprets the patterns and finds the output methodology	-Adopts the trial-and-error methodology

Figure 5. Difference between types of machine learning algorithms

3.3 Overview of Lean Six Sigma

Lean six sigma methodology was established by combining management with continuous improvement theories. In the 1980s, lean six sigma programs were first introduced by Motorola; then, it was utilized with various lean methodologies in the 2000s (Snee 2010). According to Laureani and Antony (2012) , lean six sigma is a methodology that combines lean and six sigma principles and tools; it aims for continuous improvement to maximize the value by increasing the quality and user satisfaction and decreasing costs. The lean six sigma methodology can be used as a continuous improvement strategy for the manufacturing and service sectors. It is a set of principles which aims for dynamic continuous improvement. The integration of lean and six sigma helps organizations achieve their goals by being more effective and efficient (Albliwi et al. 2015). The implementation of lean six sigma in the manufacturing sector leads to many benefits for the organizations, such as: improving the customer's satisfaction, quality levels and the productivity, reducing the inventory, cost, cycle and lead time, despite the benefits it has limitations summarized by the following there are limited number of implemented case studies of lean six sigma in manufacturing sectors. Lack of certification standards for lean six sigma. Lack of experienced experts. Hence, organizations must overcome these limitations and challenges to reach the continuous improvement level for implementing lean six sigma in the manufacturing sector. As for the service sector, implementing lean six sigma can improve the operations, service quality, cycle and lead time, employee satisfaction, and reduce the error rate, cost, delay, and complaint (Singh and Rathi 2018). Cusumano et al. (2021) defined lean in official statistics as eliminating waste by reducing variability associated with user information, provider information, and statistical processes. The utilization of lean six sigma enables the elimination of non-value adding activities while increasing the quality and decreasing the cost (Auksztol 2021). Lean six sigma has many tools and techniques, histogram, pareto, control charts, and fishbone diagram are examples of lean six sigma tools. One the other hand, quality function deployment, experiment design, statistical process control and failure mode and effects analysis are examples of lean six sigma techniques. The main difference between the tools and techniques is that the tools are used for a narrow focus while the techniques are used for a wide application (Uluskan 2019).

3.4 Role of Big Data Analytics in Official Statistics

The UN Statistical Commission provided a strategic vision for incorporating big data analytics with official statistics while focusing on finding solutions for its associated challenges (UN Committee of Experts on Big Data and Data Science for Official Statistics 2020). According to Horobets (2021), one of the outputs of the "Collaboration in Research and Methodology for Official Statistics" conference in 2014 is identifying big data as a critical source for producing official statistics. Big data is an infinite flow of unstructured, irregular, and uncontrolled information. This flow of information led to identifying new data sources that can be used by National Statistics Organizations. The role of big data analytics in official statistics is to have alternative data sources due to the survey's high costs and non-responded level in conventional statistics. Choosing suitable big data analytics tools, techniques, and data sources for producing official statistics is necessary to reduce total costs (Uluwiyah 2016). Additionally, Daas et al. (2015) discussed that due to the decreasing response rate, National Statistics Organizations began to utilize big data as an

alternative data source to supplement or replace the conventional statistical processes. Many National Statistics Organizations employed big data with official statistics. Statistics of Netherlands utilized R software to analyze traffic loops and social media messages. They resulted in gathering more detailed statistics while reducing the time and budget. But employing big data with official statistics caused some challenges such as missing data, data volatility and legal issues. Statistics of Indonesia experienced efficient data collection by using Mobile Positioning Data to set Indonesia metropolitans' areas. Moreover, they used web scraping to collect tourism statistics, consumer price index, and unemployment rates. Statistics of Indonesia faced challenges related to data quality, accessibility, and skillset (Noviyanti et al. 2020). The Australian Bureau of Statistics increased the response rate by using big data analytics, survey sample, and finite population for the Australian Agricultural Census, during the implementation, they faced issues related to missing data and measurement errors (Kim and Tam 2021). A case study was implemented for the U.S consumer price index by using web scraping and the available online data to reduce cost and time (Harchaoui and Janssen 2018). Web scraping is one of big data's collection sources. It accelerates the data collection process flexibly, reducing the implementation costs. There are many applications for web scraping to collect more frequent data in less time and cost. An application of collecting and analyzing the poverty indicators using web scraped data from World Bank Poverty (Commission 2014). Another application of calculating the Italian Harmonized Index of Consumer Price (HICP) by web scraping the consumer prices. Also applying web scraping to forecast the products prices levels (Polidoro et al. 2015). Moreover, employing web scraping to calculate the Norwegian consumer price index (Krsinich 2016).

3.5 Role of Lean Six Sigma in Official Statistics

A Polish Statistician applied lean tools in a survey project; first, they identified the stakeholders' needs, then they determined the processes required to be modified by developing the value stream. After that, they reduced the data scope based on eliminating redundant data and using data collected from another project. Thus, the number of surveyed units and the collected data increased. The Polish statistical office concluded that lean six sigma could effectively be implemented in statistical processes (Auksztol 2021). The Central Statistics Office of Ireland implemented lean tools in projects to decrease the projects' budgets. The office started with a pilot project, then initiated training programs for the employees to apply lean tools. The office then moved to the framework design stage to identify the proposed modifications. The focus was on shortening the project time and reducing the resources and errors using DMAIC lean methodology. The results of the project confirmed the effectiveness of applying lean tools in official statistics production (McSweeney and Moore 2015).Moreover, another study carried out in the Netherlands incorporated lean six sigma tools in the operational processes and quality policies of the statistical Netherlands office by using the DMAIC methodology, which eventually led to the continual improvement of the methods (Smekens and Zeelenberg 2015). On a similar note, the official statistics office of Canada integrated lean six sigma tools with quality management in their official statistics production processes. This resulted in a 5% operational budget reduction (McSweeney and Moore 2015).

4. Key Research Challenges

After the analysis of the literature, various previous studies discussed the role of big data in official statistics, where several National Statistics Organizations benefited from it by reducing the processing time and the total costs. Many papers outlined the applications of big data in lean six sigma which supported the process improvement. Additionally, some papers deliberated the uses of lean six sigma in official statistics for process continual improvements. Nevertheless, the integration of official statistics with big data analytics and official statistics with lean six sigma resulted in rising key challenges. This section highlights two research challenges categories and the proposed solutions.

Research challenge category #1

Incorporating Big Data Analytics into official statistics

Integrating official statistics with big data analytics can result in several potential challenges. National Statistics Organizations find it hard to obtain meaningful information from unstructured data (Daas et al. 2015). According to Zicari (2014), significant data challenges are categorized into the process, data, and management challenges. One of the challenges that some organizations may face is the lack of the infrastructure needed to gain big data analytics (Alsghaier et al. 2017). Hariri et al. (2019) stated that IT infrastructure could cause a big challenge for using big data analytics for information processing. In addition, organizations must monitor data quality, retention, integration, security, and privacy. Furthermore, Maple (2017) discussed the privacy and security challenges for society and the government's role in shaping laws and policies for this development. Ethics is very important in big data

analytics; one of the issues that individuals face is utilizing their personal information and data commercially without their permission (Ogbuke et al. 2022). According to the current literature, the challenges in performing big data analytics are associated with security issues, privacy issues, ethical issues, lack of laws and regulations for preserving individual information, and inadequate infrastructure and IT infrastructure. Based on the above discussion, the following challenges must be addressed and resolved.

Lack of infrastructure is a significant challenge while implementing big data with official statistics. Using large datasets requires an advanced infrastructure with proper data processing, data warehouse, and data analysis technologies. National Statistics Organizations must invest in providing the right technologies and infrastructure to improve statistical processes and outputs. Also, data quality is considered a critical challenge to address due to the dependency of machine learning algorithms on big data sources. Failure to meet the data quality requirements can lead to unsatisfied results, which will affect the organization's reputation. A monitoring and fixing system should be implemented continuously to overcome this issue. Moreover, data privacy is another main challenge for applying big data in official statistics, even though privacy is vital for National statistics Organizations. Obtaining data without the consent of the users can result in potential legal threads. While implementing big data analytics with official statistics, transparency, obtaining consent, information security, accountability, and preserving individuals' rights should be considered. In addition to that, publishing laws for protecting the user's personal information should be considered.

Research challenge category #2

Incorporating Lean Six Sigma in official statistics

According to McSweeney and Moore (2015), implanting lean six sigma with official statistics in National Statistics Organizations entails some challenges. A hierarchy organization structure is a significant challenge for integrating lean six sigma with official statistics, which can be a blockage for applying changes. National Statistics Organizations must adopt a more flexible organizational structure to enhance communication, reduce waste, and eliminate non-added value activities. Lack of management support is a critical challenge that must be addressed by engaging leaders with the employees, encouraging changes in culture, and enforcing leadership support. Also, the lack of a common goal is a challenge that can lead to difficulties in adopting changes and losing the focus to improve the performance. National Statistical Organizations must identify an apparent purpose by focusing on the customers' needs and committing to achieving it.

5. Conclusion and Future Work

While facilitating the daily life activities and business processes, fast growing technology brings some challenges with itself. As it is very well-known, big data analytics can be utilized in many aspects now, including the official statistics processes. The benefits of using big data tools and lean six sigma are obvious in the literature, however, there is very limited research about a comprehensive list of the challenges. This paper identified these challenges by a systematic literature review technique. It appeared that the most critical challenges for integrating big data analytics with official statistics are lack of proper infrastructure, data quality, and data privacy. The most critical challenges for integrating official statistics with lean six sigma are the hierarchical organization structure, lack of management support, and lack of common goal. These findings show that the organizations need a systematic process to generate official statistics by using the advanced technological methods. Therefore, as a future work, it is suggested to develop a framework which may help organizations integrate the big data analytics and lean six sigma into the process of generating official statistics. This framework may also support organizations in overcoming the stated challenges in this paper and maximize the benefits to produce more efficient and timely statistics.

References

Adolph, M., Big Data, Its Enablers and Standards, *PIK-Praxis Der Informationsverarbeitung Und Kommunikation*, vol. 37, no. 3, pp. 197–204, 2014.

Albliwi, S. A., Antony, J. and halim Lim, S. A., A Systematic Review of Lean Six Sigma for the Manufacturing Industry, *Business Process Management Journal*, 2015.

Alsghaier, H., Akour, M., Shehabat, I. and Aldiabat, S., The Importance of Big Data Analytics in Business: A Case Study, *American Journal of Software Engineering and Applications*, vol. 6, no. 4, pp. 111–15, 2017.

Alwan, H. B., and Ku-Mahamud, K. R., Big Data: Definition, Characteristics, Life Cycle, Applications, and Challenges, *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, vol. 769, no. 1, p. 012007, 2020.

Auksztol, J., Lean Official Statistics Concept Based on the Working Conditions Survey, *Wiadomości Statystyczne. The Polish Statistician*, vol. 66, no. 12, pp. 75–97, 2021.

Chawda, R. K., and Thakur, G., Big Data and Advanced Analytics Tools, *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*, IEEE, pp. 1–8, 2016.

Commission, U. N. S., Big Data and Modernization of Statistical Systems, *Report of the Secretary-General E/CN*, vol. 3, no. 11, 2014.

Cox, M., and Ellsworth, D., Managing Big Data for Scientific Visualization, *ACM Siggraph*, vol. 97, no. 1, pp. 21–38, 1997.

Cully, A., Clune, J., Tarapore, D. and Mouret, J.-B., Robots That Can Adapt like Animals, *Nature*, vol. 521, no. 7553, pp. 503–7, 2015.

Cusumano, M. A., Holweg, M., Howell, J., Netland, T., Shah, R., Shook, J., Ward, P. and Womack, J., Commentaries on "The Lenses of Lean," *Journal of Operations Management*, vol. 67, no. 5, pp. 627–39, 2021.

Daas, P. J. H., Puts, M. J., Buelens, B. and Hurk, P. A. M. van den, Big Data as a Source for Official Statistics, *Journal of Official Statistics*, vol. 31, no. 2, p. 249, 2015.

Eck, N. J. van and Waltman, L., Citation-Based Clustering of Publications Using CitNetExplorer and VOSviewer, *Scientometrics*, vol. 111, no. 2, pp. 1053–70, 2017.

Frizzo-Barker, J., Chow-White, P. A., Mozafari, M. and Ha, D., An Empirical Study of the Rise of Big Data in Business Scholarship, *International Journal of Information Management*, vol. 36, no. 3, pp. 403–13, 2016a.

Frizzo-Barker, J., Chow-White, P. A., Mozafari, M. and Ha, D., An Empirical Study of the Rise of Big Data in Business Scholarship, *International Journal of Information Management*, vol. 36, no. 3, pp. 403–13, 2016b.

Goundar, S., Bhardwaj, A., Singh, S., Singh, M. and Gururaj, H. L., Big Data and Big Data Analytics: A Review of Tools and Its Application, *Applications of Big Data in Large-and Small-Scale Systems*, pp. 1–19, 2021.

Gupta, B., Kumar, A., and Dwivedi, R. K., Big Data and Its Applications– A Review, *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pp. 146–49, 2018.

Harchaoui, T. M. and Janssen, R. v, How Can Big Data Enhance the Timeliness of Official Statistics?: The Case of the US Consumer Price Index, *International Journal of Forecasting*, vol. 34, no. 2, pp. 225–34, 2018.

Hariri, R. H., Fredericks, E. M. and Bowers, K. M., Uncertainty in Big Data Analytics: Survey, Opportunities, and Challenges, *Journal of Big Data*, vol. 6, no. 1, pp. 1–16, 2019.

Horobets, O., Research Data as a Result of Research Activities: The Role and Significance for the Official Statistics, *Journal of the Knowledge Economy*, vol. 12, no. 3, pp. 1424–36, 2021.

Kim, J. and Tam, S., Data Integration by Combining Big Data and Survey Sample Data for Finite Population Inference, *International Statistical Review*, vol. 89, no. 2, pp. 382–401, 2021.

Kitchin, R. and McArdle, G., What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets, *Big Data & Society*, vol. 3, no. 1, p. 2053951716631130, 2016.

Krsinich, F., The FEWS Index: Fixed Effects with a Window Splice., *Journal of Official Statistics (JOS)*, vol. 32, no. 2, 2016.

Laureani, A. and Antony, J., Critical Success Factors for the Effective Implementation of Lean Sigma: Results from an Empirical Study and Agenda for Future Research, *International Journal of Lean Six Sigma*, 2012.

Maple, C., Security and Privacy in the Internet of Things, *Journal of Cyber Policy*, vol. 2, no. 2, pp. 155–84, 2017.

McSweeney, K. and Moore, K., Innovating to Do More with Less-the Story of Lean Six Sigma in the Central Statistics Office, Ireland, *Statistical Journal of the IAOS*, vol. 31, no. 4, pp. 587–92, 2015.

Noviyanti, I., Prabawa, P. D., Sari, D. P., Koswara, A., Lestari, T. K., Fahyuananto, M. H. and Setiawan, E., 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*, vol. 36, no. 4, pp. 943–54, 2020.

Nti, I. K., Quarcoo, J. A., Aning, J. and Fosu, G. K., A Mini-Review of Machine Learning in Big Data Analytics: Applications, Challenges, and Prospects, *Big Data Mining and Analytics*, vol. 5, no. 2, pp. 81–97, 2022.

Ogbuke, N. J., Yusuf, Y. Y., Dharma, K. and Mercangoz, B. A., Big Data Supply Chain Analytics: Ethical, Privacy and Security Challenges Posed to Business, Industries and Society, *Production Planning & Control*, vol. 33, no. 2–3, pp. 123–37, 2022.

Polidoro, F., Giannini, R., Conte, R. lo, Mosca, S. and Rossetti, F., Web Scraping Techniques to Collect Data on Consumer Electronics and Airfares for Italian HICP Compilation, *Statistical Journal of the IAOS*, vol. 31, no. 2, pp. 165–76, 2015.

Radermacher, W. J., *Official Statistics 4.0*, Springer, 2020.

Raja, R., Mukherjee, I. and Sarkar, B. K., A Systematic Review of Healthcare Big Data, *Scientific Programming*, vol. 2020, 2020.

Russom, P., Big Data Analytics, *TDWI Best Practices Report, Fourth Quarter*, vol. 19, no. 4, pp. 1–34, 2011.

Singh, M. and Rathi, R., A Structured Review of Lean Six Sigma in Various Industrial Sectors, *International Journal of Lean Six Sigma*, 2018.

Smekens, M. and Zeelenberg, K., Lean Six Sigma at Statistics Netherlands, *Statistical Journal of the IAOS*, vol. 31, no. 4, pp. 583–86, 2015.

Snee, R. D., Lean Six Sigma–Getting Better All the Time, *International Journal of Lean Six Sigma*, 2010.

Uluskan, M., Analysis of Lean Six Sigma Tools from a Multidimensional Perspective, *Total Quality Management & Business Excellence*, vol. 30, no. 9–10, pp. 1167–88, 2019.

Uluwiyah, A., Trusted Big Data for Official Statistics: Study Case: Statistics Indonesia (BPS), *2016 International Conference on Information Technology Systems and Innovation (ICITSI)*, IEEE, pp. 1–6, 2016.

UN Committee of Experts on Big Data and Data Science for Official Statistics, 2020.

Wadhera, S., Kamra, D., Kumar, A., Jain, A., and Jain, V., A Systematic Review of Big Data Tools and Application for Developments, *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)*, IEEE, pp. 561–66, 2021.

Wang, L. and Alexander, C. A., Machine Learning in Big Data, *International Journal of Mathematical, Engineering and Management Sciences*, vol. 1, no. 2, p. 52, 2016.

Zhou, L., Pan, S., Wang, J. and Vasilakos, A. v, Machine Learning on Big Data: Opportunities and Challenges, *Neurocomputing*, vol. 237, pp. 350–61, 2017.

Zicari, R. v, Big Data: Challenges and Opportunities, *Big Data Computing*, vol. 564, p. 103, 2014.

Biography

Farah Al Jaghoub, she earned her bachelor's degree of Science in Industrial Engineering and Engineering Management from University of Sharjah in 2015. Currently she is undertaking her master's degree of Science in Engineering Management from University of Sharjah. Since January 2016, she is employed at Dubai Statistics Center as a Senior Quality Officer. Farah's research focuses on lean six sigma, industry 4.0, and big data analytics.

Zehra Canan Araci is an Assistant Professor at the department of Industrial Engineering and Engineering Management at University of Sharjah in the UAE. Dr. Zehra holds a PhD degree in Lean Product and Process Development from Cranfield University in the UK. She has 13 years of academic and industrial experience in quality management and lean systems at prestigious organizations such as Rolls-Royce Plc, Airbus, IKEA suppliers, Caltec, Arcelik, and Al-Zahra Hospital Dubai. Dr. Zehra's research focuses on continuous improvement activities, productivity, lean systems, quality management systems, and business excellence. Her recent developing research interest is in the field of Industry 4.0.

Concetta Semeraro is an assistant professor at the University of Sharjah, United Arab Emirates, in Industrial Engineering and Engineering Management. She was a post-doc at the University of Lorraine, France in 2021. She received her double PhD FIRST-CLASS HONOR in Management and Mechanical Engineering at Polytechnic University of Bari (Italy) and in Computer Science at the University of Lorraine (France) in 2020. For her PhD, she gained expertise on Industry 4.0, modelling and simulation, machine learning, cyber-physical systems, and digital twin to improve manufacturing processes' efficiency and smartness. She led several industrial projects in lean manufacturing, Industry 4.0 technologies, manufacturing execution system (MES) applications and digital twin prototypes for manufacturing processes.