Big data and statistical production in Latin America and the Caribbean: perspectives from National Statistics Offices

Andrea Diniz da Silva
(corresponding author)
Escola Nacional de Ciências Estatísticas – ENCE/IBGE, Rio de Janeiro, Brazil
Email: andrea.silva@ibge.gov.br

Elizabeth Belo Hypólito
Instituto Brasileiro de Geografia e Estatística – IBGE, Rio de Janeiro, Brazil
Email: elizabeth.hypolito@ibge.gov.br

Fabio Lucas Pimentel de Oliveira
Instituto de Pesquisa e Planejamento Urbano e Regional – IPPUR/UFRJ, Rio de Janeiro, Brazil
Email: fabio.oliveira@ippur.ufrj.br

Marcus André Alves Zimmermann Vieira
Fundação Bradesco, São Paulo, Brazil
Email: marcusazimmermann@gmail.com

Datafication is one of the primary characteristics of the epochal change that has been underway across the globe in this first quarter of the new century. This era is led by the generation of big data, with diverse and widely distributed sources in categories such as financial transactions, remote sensing, health monitoring, telemetry, consumption measurement and social networks. The use of this type of data poses challenges associated with technological, methodological and structural concerns. To monitor advances and learn about the challenges of using big data to produce statistics in Latin America and the Caribbean, the UN Regional Hub for Big Data in Brazil developed the annual series International Consultation on the Use of Big Data. The two editions of the consultation, which were conducted with national statistics offices in 2022 and 2023, demonstrated that 10 countries had used big data to produce official or experimental statistics in the previous two years. This revealed a promising scenario, particularly when considering that the countries already using big data have regional roles, motivating and supporting other countries in the use of big data.

Keywords
big data,
official statistics,
experimental statistics,
Latin America and the Caribbean

Online first publication date: 6 December 2024

Introduction

Datafication, referring to the process of transforming actions into quantifiable data, is one of the main characteristics of the epochal change that has been underway across the globe in the first quarter of the new century (Mayer-Schönberger–Cukier 2013). The corresponding digitalisation, driven by the public availability of *internet access* that began in the mid-1990s, includes the advancement of social networks, cloud computing, the development of complex algorithms and artificial intelligence. These advances have expanded the scope and scale of methods for collecting and analysing the digital traces left by connected human and non-human users (Lemos 2021).

This techno-informational stage is led by the generation of big data, whose attributes and definition (although open and subject to recurring revisions) refer to large volume, high velocity and wide variety (Macfeely 2019). Some arguments have even expanded such attributes to an exorbitant 42 terms, all beginning with the letter V, to qualify big data (Schafer 2017). Big data sources are diverse and broadly distributed across categories such as commercial transactions, remote sensing, health monitoring, telemetry, consumption measurement and social media monitoring (Glasson et al. 2013).

The ubiquitous availability of big data poses important epistemological and ethical questions for national statistics offices (NSOs) worldwide (Kitchin 2015). Considering statistical standards and principles, we can highlight the absence of standardisation and subsequent quality assurance for some sources, the underdeveloped experience and fragile technological structure of the public sector for managing these data in developing countries and outdated legal-normative apparatus, particularly those related to data ownership, privacy and anonymisation, with generation that is based on continuously innovating connected devices. These concerns have permeated reflections on the uses linked to the production of official statistics for at least a decade.

To seek strategies to address these limitations, in 2014, the United Nations founded the Committee of Experts on Big Data and Data Science (UNCEBD), which was initially called the Global Working Group and is now known as UNBIGDATA. The inventory performed by the UNCEBD, based on consultation with national and international organisations, contains more than 200 references to cases involving the use of big data for statistical production (Macfeely 2019). Halderen et al. (2021) also gathered cases from countries using big data for statistics. The authors identified 22 countries using big data to produce indicators proposed by the 2030 Agenda for Sustainable Development, including Germany, Australia, Austria, Canada, China, Colombia, Spain, Finland, France, Netherlands, India, Ireland, Italy, Japan, Mexico, New Zealand, Portugal, Sweden, Switzerland, Turkey, Ukraine and the Philippines. The latter deserves attention because it produces indicators for 15 of the 17 SDGs, such as goal 11 – sustainable cities and communities.

In effect, what is at stake is the redefinition of the scientific and political paradigm in which reality is (de)coded by data. It is no coincidence that the applicability of big data for the preparation of official statistics is one of the challenges facing NSOs in Latin America and the Caribbean. One concern is the pressure exerted on the deliberate pace of structured, clean and simple production, although it can be costly, slow and subject to non-responses in certain situations and changes in respondents' preferences derived from the research and survey methods that form the essential reputational basis for national authorities (Pfefferman 2015).

Several studies have described NSOs' efforts to introduce big data into the official statistics framework (Abdulkadri et al. 2016, Abbas et al. 2023, Abraham 2022, Braaksma–Zeelenberg 2020). The on-going experiments consider the definition, scope and sources of big data beyond the etymology of the concept, noting the emergence of a disputed ecosystem of informational assets in which government agencies, companies, financial institutions, research institutions and other entities participate. The potential mix between traditional statistics and a renewed perception regarding costs, technological processing conditions and inputs for the elaboration of big data involves web scraping, satellite imagery, mobile phone data, sensors and metrological devices connected to means of transport (aeroplanes, trains, ships, cars, buses and even bicycles) and infrastructure in general (motorways, lighting).

This conjunction requires the development of new capabilities and skills to manage data obtained in a shorter time frame to ensure the production of reliable and high-quality information protected by methods (sample nature, matching difficulties, interoperability and temporality, among other concerns) and processing devices of governance (budgets, legislation, infrastructure, privacy and degree of digital literacy). If big data is not the perfect cure for the informational deficiencies that arise from traditional statistical processes (Diaz-Bone–Horvath 2021), real-time adherence to the dynamics and emergent needs of everyday life is indisputable, giving it intrinsic value (Florescu et al. 2014).

This is evidenced by the monitoring of natural tragedies, which, for example, motivated the creation of the Rio Operations Centre in the city of Rio de Janeiro, Brazil and the dependence on social interactions connected to nowcasting, which was reinforced during the Covid-19 pandemic by checking momentarily dense places and depicted by monitoring alternative transit routes, passenger transport services and delivery times of online purchases through delivery apps.

Why monitor the use of big data by NSOs in Latin America and the Caribbean? Because statistical authorities in this region can and must participate in the global debate surrounding issues that are still pending and are an obstacle to the concrete application of these data in a substantive way to leverage the many opportunities provided. Such issues generally refer to the curation, storage, confidentiality and dissemination of big data. Solving them requires the mediating role of NSOs

concerning the arrangements among governments, markets and society intended to ensure the functioning and permanent supply of mass public data.

This is a debate that encompasses at least four types of problems for Latin America and the Caribbean, including (i) geopolitical, given the informational asymmetry related to the platforms and foundational structures of surveillance capitalism (Srnicek 2017, Zuboff 2021); (ii) socioeconomic, considering the inequalities and incipient digital illiteracy in the countries of the region that inhibit the democratisation of access to tools such as artificial intelligence;[1] (iii) political, in the form of the aforementioned dimensions of privacy, legislation, governance and public financing; and (iv) technical, involving the difficulties of using big data in contrast to the traditional processes of producing official statistics.

This study emphasises the previously introduced fourth dimension, presenting a reflection specifically focused on the recent experiences of regional NSOs using big data. This choice reflects a deliberate selection for the national scale of analysis since these institutions are the statistical authorities of their respective countries. However, this does not imply that experiences at the subnational level are absent, including those led by private corporations and third-sector institutions. While these experiences should not be overlooked, they fall beyond the scope of this analysis and are not the focus of this research.

In any case, it is urgent to have reliable and accessible data produced quickly and disaggregated thematically and geographically. The more solid the legal, institutional and financial apparatus under which these statistics are prepared is, the greater the value of big data will be for public and private agents on national and subnational scales. Therefore, conducting consultations with NSOs to monitor the existence of projects that use big data and/or future initiatives that would eventually consider this use in official research is essential, contributing to the empowerment and emancipation of society by making information of this type publicly available.

First and second consultations: method and responding countries

To monitor NSOs' use of big data to produce statistics in Latin America and the Caribbean, the UN Regional Hub for Big Data in Brazil, with the support of the United Nations Economic Commission for Latin America and the Caribbean, began an annual series of consultations with NSOs. Consultations were distributed to all

[1] For a diagnosis, see the Latin American Artificial Intelligence Index, which is published by the National Centre for Artificial Intelligence of Chile in partnership with the Economic Commission for Latin America and the Caribbean, available at https://indicelatam.cl/ (downloaded on March 2024). Another initiative that deserves mention is the project Artificial Intelligence and Public Policies (IAPP), which is a collaborative initiative of the Centre for Studies Applied to the Public Sector of the Federal University of Goiás, the Centre for Public Policy Studies at the State University of Campinas and the National School of Statistical Sciences of the Brazilian Institute of Geography and Statistics, available at https://ciap.org.br/ (downloaded on March 2024).

countries in Latin America and the Caribbean in the first quarter of 2022 and 2023. The results of the first consultation were published by Silva et al. (2023).

As in the previous year, the questionnaire applied in 2023 included questions to identify the country, NSO and the respondent, all with mandatory answers, in addition to closed questions on the topic and open questions to identify advances and challenges related to the use of big data to produce statistics. The investigation of the main topic, the use of big data, was encapsulated in 12 closed questions.

Four main questions were posed to identify the use of big data by each responding NSO (translated here from Portuguese), including 'Does the statistics office use big data sources in the production of official statistics?', 'Does the statistics office use big data sources in production of experimental statistics?', 'Is the statistics office conducting studies or tests for the use of big data sources in the production of statistics?' and 'Is the statistics office considering conducting studies or tests for the use of big data sources in the production of statistics?'. The answer options are binary (Yes or No).

In cases where the answer was 'Yes', additional mandatory questions were asked about the type of big data and the topics covered. The answer options were adapted from MacFeely (2019) in the original language and translated into Portuguese and Spanish. The response options for the type of big data included web scraping, satellite imagery, mobile phone data, social network records, scanned data, consumption metres, financial transactions, road sensors, health records and ship tracking. The answer options for the topics included prices, population, mobility, agriculture, the labour market, health, poverty and inequality and disaster risk reduction.

Additionally, four open questions were asked to understand the available resources and regional needs, including 'Does the statistics office have any resources or expertise in the use of big data that can be shared with other countries?', 'What are the limitations and needs for the use of big data in statistical production in your country?', 'Indicate the training topics of greatest interest or immediate need' and 'Are there any other contributions that could help the Hub plan actions to support the use of big data in the region?'.

The full questionnaire is available on request.

Coverage

In 2023, there was an increase in the response rate from NSOs in the Caribbean, rising from 2 to 6 of the 28 countries in the region between 2022 and 2023. However, the number of Central and South American respondents decreases. Of the 24 countries in that region, 14 responded in 2022, while only 11 responded in 2023 (Figure 1). The final balance was positive, increasing from 16 to 17 responding countries in 2023.

Figure 1

Countries in Latin America and the Caribbean that responded to the consultations

Source: UN Regional Hub for Big Data in Brazil. International Consultation on the Use of Big Data for Statistics in Latin America and the Caribbean.

The low relative participation rate of Caribbean countries may be attributable to the lack of progress in this area in the region, as previously noted by Abdulkadri et al. (2016). The authors reference a consultation conducted by the United Nations Economic Commission for Latin America and the Caribbean with 23 countries, of which only 10 responded, with 7 of them reporting that they were not aware of any initiative involving the use of big data for statistical production.

The use of big data for statistical production

In this section, we analyse the responses from the 11 countries that responded to both consultations (i.e. the 2022 and 2023 rounds). As mentioned previously, in 2022, 16 countries responded to the first International Consultation on the Use of Big Data in Latin America and the Caribbean. Five countries that responded to the consultation in 2022 did not respond in 2023; however, six new countries responded in 2023, totalling 17 participating countries in 2023. The 11 countries that participated in the two rounds of consultation are presented in Figure 2. It should be noted that seven of these countries are South American.

Figure 2

Countries that responded to the first and second International Consultation on the Use of Big Data in Latin America and the Caribbean, 2022 and 2023

Source: UN Regional Hub for Big Data in Brazil. International Consultation on the Use of Big Data for Statistics in Latin America and the Caribbean.

Six of the 11 countries reported using big data for official statistics in 2022: Argentina, Brazil, Colombia, Costa Rica, Peru and Uruguay. Four of them, Argentina, Brazil, Colombia and Peru, reported using big data in 2023. In addition, Ecuador and Paraguay reported using data sources. Five countries used big data for experimental statistics: Argentina, Brazil, Colombia, Mexico and Uruguay. Peru and the Dominican Republic also reported using big data in 2023 (Table 1). Of the 11 countries that responded to the consultations in 2022 and 2023, only Cuba did not use big data for official or experimental statistics.

Ten countries reported conducting studies or tests using big data in 2022: Argentina, Brazil, Colombia, Costa Rica, Cuba, Ecuador, Mexico, Peru, the Dominican Republic and Uruguay. Of these, the Dominican Republic was the only country to answer no to the same question in 2023. However, notably, this country responded that it was using big data for experimental statistics in 2023 and was not doing so in 2022. Thus, a transition from studying to producing experimental statistics may have occurred. Furthermore, the new entry, Paraguay, reported conducting studies using big data in 2023 (Table 1).

Table 1

Number of countries that responded to the International Consultation on the Use of Big Data, total and purpose of use

Purpose of use	2022	2023
Respondents	11	11
Use big data for official statistics	6	6
Use big data for experimental statistics	5	7
Are conducting studies or tests on the use of big data	10	10
Are considering using big data	1	2

Source: UN Regional Hub for Big Data in Brazil. International Consultation on the Use of Big Data for Statistics in Latin America and the Caribbean.

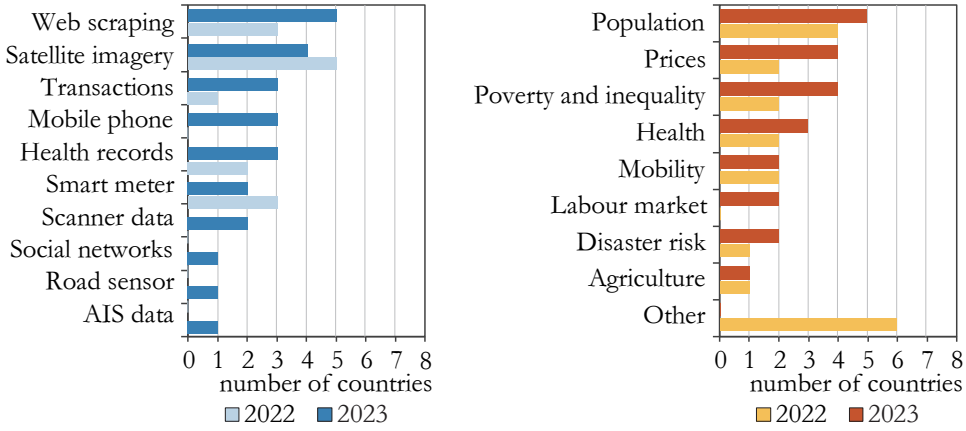
Web scraping, used in five countries, was the most common source for producing official statistics in 2023, followed by satellite imagery in four countries and financial transaction data, mobile phone data and health records in three countries. Furthermore, an increase in the types of sources used occurred in 2023 compared with the previous year, with the use of mobile phone data, social networks, road sensors and ship tracking reported for the first time in 2023 (Figure 3).

Population statistics was the topic for which more countries (five) used big data in official statistics in 2023, followed by prices, poverty and inequality, reported by four countries. Furthermore, the number of countries increased for practically all topics compared to 2022. The reduction in the number of countries that marked the other option in the questionnaire is noteworthy, from six in 2022 to zero in 2023. This difference may be attributable to changes in the questionnaire to improve the response options provided to respondents (Figure 3).

Figure 3 shows that web scraping was also the most used source of big data for the production of experimental statistics in 2023, reported by seven countries, five more than in the previous year. Social networks and satellite imagery, used by four countries, are tied as the second most common source. These sources also showed an increase in the number of countries using it compared with 2022. It is also noteworthy that road sensor data and ship tracking began to be used, while mobile phone data was no longer used. In the latter case, experimental statistics may have become official.

Figure 3

Number of countries that reported using big data for official statistics according to the type of big data and the topic for which it was used, Latin America and the Caribbean

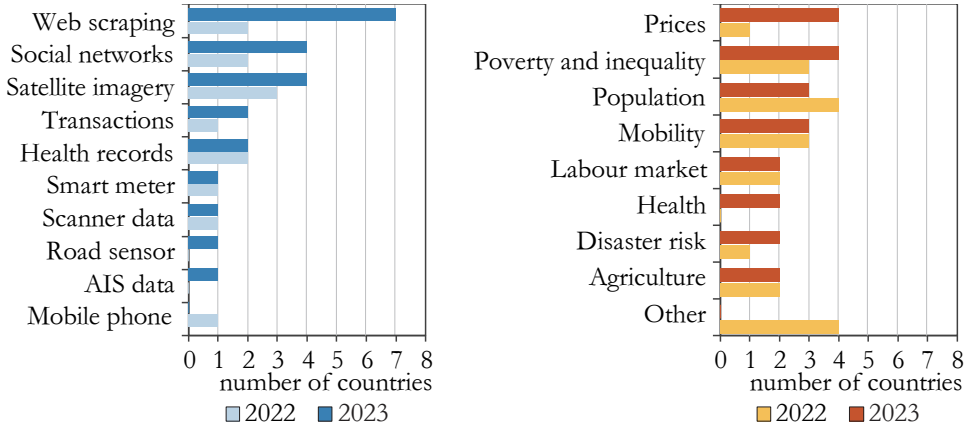


Source: UN Regional Hub for Big Data in Brazil. International Consultation on the Use of Big Data for Statistics in Latin America and the Caribbean.

Regarding the themes for which big data was used in the production of experimental statistics, prices as well as poverty and inequality stood out in four countries, followed by population and mobility in three countries. Health, which did not appear in the previous year, became a theme for two countries in 2023.

Figure 4

Number of countries using big data for experimental statistics, according to the type of big data and the topic for which it was used, Latin America and the Caribbean



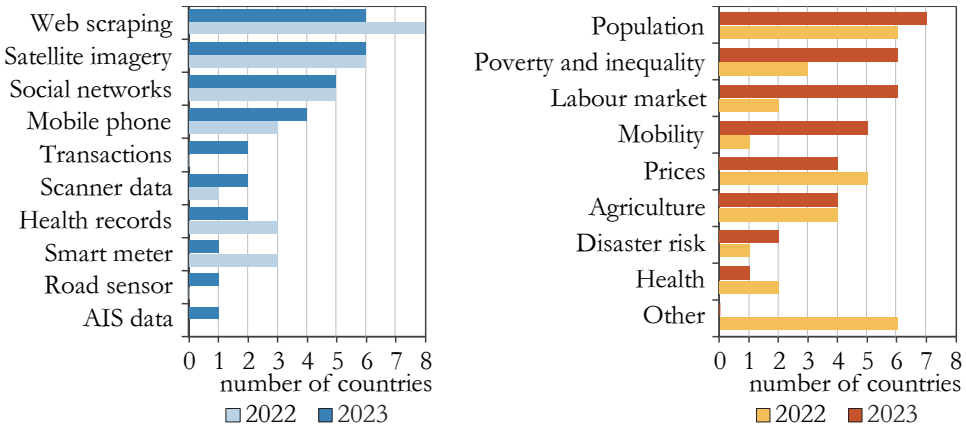
Source: UN Regional Hub for Big Data in Brazil. International Consultation on the Use of Big Data for Statistics in Latin America and the Caribbean.

Figure 4 shows that web scraping and satellite imagery were also the most studied big data sources in 2023, being reported by six countries, followed by mobile phone data in five countries and social network records in four. In relation to the previous year, three new sources, data from financial transactions, road sensors and ship tracking, began to be studied.

The most studied themes were population, studied by seven countries; poverty and inequality and labour market, by six countries; and mobility, by five countries. These last two themes saw the largest increase in the number of countries compared to 2022.

Figure 5

Number of countries studying the use of big data to produce statistics, according to the type of big data and the topic for which it was used, Latin America and the Caribbean



Source: UN Regional Hub for Big Data in Brazil. International Consultation on the Use of Big Data for Statistics in Latin America and the Caribbean.

Regional needs and resources

To understand NSOs' needs and identify the available resources in Latin America and the Caribbean for possible regional cooperation, two questions were asked, including 'What are the limitations and needs for the use of big data for statistical production in your country?' and 'Does the statistics office have some resources or expertise in the use of big data that can be shared with other countries?'.

The main limitation reported was the lack of skilled labour to use these new data sources, particularly regarding the use of technologies and methods. Concerns were also highlighted regarding the lack of a legal framework to support the use of this type of data, both in terms of guaranteeing access and privacy. The lack of technological infrastructure, especially space for processing and storing data, was also a concern for NSOs, as was the lack of financial resources in general. One NSO noted the country's

small population size as limitation to using big data, a factor that is shared by several countries, especially in the Caribbean.

Seven countries reported having some resources or expertise in the use of big data that could be shared with other countries, with the most cited expertise was the use of web scraping and satellite imagery. However, methods for data analysis and artificial intelligence were also mentioned.

The consultation includes a question about the training of greatest interest or immediate need for NSOs, which will help in planning future workshops and webinars to be offered by the Hub. Topics such as modernisation of the statistical system, machine learning, artificial intelligence, natural language processing, manipulation of satellite imagery and use of specific tools such as Python were highlighted.

Finally, NSOs were invited to make suggestions about any other contributions that could help the Hub plan actions to support the use of big data in the region. The most common suggestions were offering training and creating space to share experiences. Technological infrastructure, agreements, consultancy and financing were also among the contributions considered welcome by the NSOs that responded to the consultation.

Final considerations

Among the countries of Latin America and the Caribbean, the scope and use of big data encounter serious challenges related to the socioeconomic, political and institutional characteristics that define the region. While these characteristics were not the focus of the analysis in this article, they shape the conditions within which the efforts of NSOs in the region unfold.

In the past few years, advances have been made in the use of big data for statistical production by NSOs in Latin America and the Caribbean, particularly in forums organised by the United Nations. The UN Statistical Commission and the Statistical Conference of the Americas have been the stage for countless examples of countries producing official and experimental statistics using big data, either exclusively or in conjunction with data from administrative records or traditional surveys. More than a dozen success stories in the region were presented in a series of webinars organised by the UN Regional Hub for Big Data in Brazil in 2023.

The two editions of the International Consultation on the Use of Big Data conducted with NSOs in 2022 and 2023 showed that 10 countries have used big data to produce official or experimental statistics in the last two years. This reveals a promising scenario, particularly considering that countries that already use big data have regional influence, motivating and supporting other countries in the use of this type of data.

Despite advances, challenges associated with NSOs' precarious technological infrastructure, the scarcity of qualified human resources, obsolete legislation and the need for methods that can produce quality statistics with this (not so) new class were identified. Such challenges are not exclusive to the use of big data. Insufficient infrastructure and personnel appear in the most diverse contexts, whether associated with the use of alternative or traditional data. This finding is predominant in large-scale traditional operations such as population and agricultural censuses (Skinner 2018, O'Hare 2018). Likewise, the obsolescence of legislation regarding respondents' compulsory cooperation is a problem when collecting personal and/or institutional data and access to administrative records.

Big data still has a long way to go in the methodological field compared to traditional research methods. The first census, whose data still exist and are considered to have good quality, dates back to the 1st century, while the development of the basis for modern sampling, the confidence interval that is attributed to Jerzy Neyman dates back to 1937. Therefore, the use of traditional data has been in place for over 80 years (Timeline 2013). Although the development of methods occurs at a faster pace in contemporary times, a suitable consolidated theory that supports the use of big data has not yet been developed.

Alternatives that consider big data collection as a non-probability sampling method have led to the use of sampling theory for this type of data (Golini–Righi 2024). The use of inverse sampling and the integration of big data with probabilistic survey data are alternatives that have been found to reduce selection bias. The use of multiple sources, combining administrative records and big data, is another approach that has shown positive results (Kim–Wang 2019). Although these initiatives do not yet conform to a formal theory on the use of big data and do not yet place big data under the umbrella of sampling, they do point out ways to obtain benefits from this alternative data source.

The upcoming edition of the consultation, which will be conducted in 2024, will include in-depth interviews with the countries that participated in at least one of the rounds to allow further in-depth analysis. It will include a retrospective on 19 countries and a prospective on the participants in the third round. It is expected that the use of big data has become real in a larger range of countries in the region.

REFERENCES

ABBAS, S.W.–HAMID, M.–ALKANHEL , R.–ABDALLAH, H. A. (2023): Official statistics and big data processing with artificial intelligence: capacity indicators for public sector organizations *Systems* 11 (8): 424. https://doi.org/10.3390/systems11080424

ABDULKADRI, A.–EVANS, A.–ASH, T. (2016): An assessment of big data for official statistics in the Caribbean: challenges and opportunities In: ECLAC: *Studies and Perspectives Series – The Caribbean – No. 48*. United Nations.

ABRAHAM, K. G. (2022): Big data and official statistics *Review of Income and Wealth* 68 (4): 835–861. https://doi.org/10.1111/roiw.12617

BRAAKSMA, B.–ZEELENBERG, K. (2020): Big data in official statistics *Discussion paper* Centraal Bureau voor de Statistiek, Den Haag.

DA SILVA, A. D.–MENEZES MARQUES DE OLIVEIRA, B.–GONÇALVES PEIXOTO, Í.–BRAGA SALES DE SOUZA, L. (2023): Overview of the use of big data for official statistics in Latin America and the Caribbean *Statistical Journal of the IAOS* 39 (1): 171–177. https://doi.org/10.3233/SJI-220092

DIAZ-BONE, R.–HORVATH, K. (2021): Official statistics, big data and civil society. Introducing the approach of "economics of convention" for understanding the rise of new data worlds and their implications *Statistical Journal of the IAOS* 37 (1): 219–228. https://doi.org/10.3233/SJI-200733

FLORESCU, D.–KARLBERG, M.–REIS, F.–CASTILLO, P. R.–SKALIOTIS , M.–WIRTHMANNN , A. (2014): *Will 'big data' transform official statistics?* Eurostat, Luxembourg.

GLASSON, M.–TREPANIER, J.–PATRUNO, V.–DAAS, P.–SKALIOTIS , M.–KHAN, A. (2013): *What does "big data" mean for official statistics?* United Nations Economic Commission for Europe, Conference of European Statisticians.

GOLINI, N.–RIGHI, P. (2024): Integrating probability and big non-probability samples data to produce official statistics *Statistical Methods & Applications* 33: 555–580. https://doi.org/10.1007/s10260-023-00740-y

KIM, J. K.–WANG, Z. (2019): Sampling techniques for big data analysis *International Statistical Review* 87 (SI): S177–S191. https://doi.org/10.1111/insr.12290

LEMOS, A. (2021): Datafication of life *Civitas: Revista De Ciências Sociais* 21 (2): 193–202. https://doi.org/10.15448/1984-7289.2021.2.39638

MACFEELY, S. (2019): *The big (data) bang: opportunities and challenges for compiling SDG indicators* United Nations Conference on Trade and Development. https://doi.org/10.1111/1758-5899.12595

MAYER-SCHÖNBERGER, V.–CUKIER, K. (2013): *Big data. A revolution that will transform how we live, work, and think* Houghton Mifflin Harcourt, Boston, MA.

NEYMAN, J. (1937): Outline of a theory of statistical estimation based on the classical theory of probability *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 236 (767): 333–380. https://doi.org/10.1098/rsta.1937.0005

O'HARE, W. P. (2018): *2020 Census faces challenges in rural America* The Carsey School of Public Policy at the Scholars' Repository. 330. https://scholars.unh.edu/carsey/330

PFEFFERMAN, D. (2015): Methodological issues and challenges in the production of official statistics: 24th annual Morris Hansen lecture *Journal of Survey Statistics and Methodology* 3 (4): 425–483. https://doi.org/10.1093/jssam/smv035

SKINNER, C. (2018): Issues and challenges in census taking *Annual Review of Statistics and its Application* 5: 49–63. https://doi.org/10.1146/annurev-statistics-041715-033713

SRNICEK, N. (2017): *Platform capitalism* Polity Press, Cambridge.
Timeline of Statistics *Significance* 10: (6): 23–26. https://doi.org/10.1111/j.1740-9713.2013.00707.x

ZUBOFF, S. (2019): *The age of surveillance capitalism: the fight for a human future at the new frontier of power* Profile Books Ltd.

INTERNET SOURCES

HALDEREN , G. V.–BERNAL, I.–SEJERSEN, T.–JANSEN, R.–PLOUG, N. TRUSZCZYNSKI, M. (2021): Big Data for the SDGs. Country examples in compiling SDG big data para a produção estatística indicators using non-traditional data sources *Working Paper Series* ESCAP Statistics Division. SD/WP/12. https://www.unescap.org/sites/default/d8files/knowledge-products/SD_Working_Paper_no12_Jan2021_Big_data_for_SDG_indicators.pdf (downloaded: October 2024)

KITCHIN, R. (2015): *What does big data mean for official statistics? Discover Society.* https://archive.discoversociety.org/2015/07/30/what-does-big-data-mean-for-official-statistics/ (downloaded: February 2024).

SCHAFER, T. (2017): *The 42 V's of big data and data science* KD Nuggets. https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html (downloaded: February 2024)

TIMELINE (2013): *American Statistical Association and Royal Statistical Society.* https://archive.org/details/timelineof-statistics (downloaded: February 2024)