Big data and management of municipal solid waste

Big data e gestão de resíduos sólidos municipais

Andrea Diniz da Silva 1
Elizabeth Belo Hypólito 2
Fábio Lucas Pimentel de Oliveira 3
Calvin Macedo Ribeiro Borges 4
Dimitrio dos Santos Gregório 5
Fernanda Castilho Gomes de Oliveira 6
Laira Zopellaro Machado Miranda Ferreira 7

Abstract

In Brazil, the collection and management of urban solid waste is the responsibility of municipal governments. In most cases this is done on an informational basis considering population numbers to allocate resources for this purpose. Without having data capable of

1 PhD in Population, Territory and Public Statistics, Escola Nacional de Ciências Estatísticas (ENCE) - Instituto Brasileiro de Geografia e Estatística (IBGE), R. André Cavalcanti, 106, Centro, Rio de Janeiro - RJ, CEP: 20231-050. E-mail: andrea.silva@ibge.gov.br Orcid: https://orcid.org/0000-0001-9116-0162
2 PhD in Population, Territory and Public Statistics, Escola Nacional de Ciências Estatísticas (ENCE) - Instituto Brasileiro de Geografia e Estatística (IBGE), R. André Cavalcanti, 106, Centro, Rio de Janeiro - RJ, CEP: 20231-050. E-mail: elizabeth.hypolito@ibge.gov.br Orcid: https://orcid.org/0000-0003-2939-6662
3 PhD in Economic Development, Instituto de Pesquisa e Planejamento Urbano e Regional da Universidade Federal do Rio de Janeiro (IPPUR – UFRJ), Sala João do Rio - Faculdade de Letras, RJ. E-mail: fabio.oliveira@ippur.ufrj.br Orcid: https://orcid.org/0000-0002-6210-7394
4 Undergraduate Student of Statistics, Escola Nacional de Ciências Estatísticas (ENCE) - Instituto Brasileiro de Geografia e Estatística (IBGE), R. André Cavalcanti, 106, Centro, Rio de Janeiro - RJ, CEP: 20231-050. E-mail: calvinmacedo@gmail.com Orcid: https://orcid.org/0009-0006-8149-308X
5 Master's Student in Population, Territory and Public Statistics, Escola Nacional de Ciências Estatísticas (ENCE) - Instituto Brasileiro de Geografia e Estatística (IBGE), R. André Cavalcanti, 106, Centro, Rio de Janeiro - RJ, CEP: 20231-050. E-mail: dimitrio.gregorio@gmail.com Orcid: https://orcid.org/0000-0003-4556-3344
6 Undergraduate Student of Statistics, Escola Nacional de Ciências Estatísticas (ENCE) - Instituto Brasileiro de Geografia e Estatística (IBGE), R. André Cavalcanti, 106, Centro, Rio de Janeiro - RJ, CEP: 20231-050. E-mail: fernanda03castilho@gmail.com Orcid: https://orcid.org/0009-0001-7105-7958
7 Undergraduate Student of Statistics, Escola Nacional de Ciências Estatísticas (ENCE) - Instituto Brasileiro de Geografia e Estatística (IBGE), R. André Cavalcanti, 106, Centro, Rio de Janeiro - RJ, CEP: 20231-050. E-mail: lairazop@gmail.com Orcid: https://orcid.org/0009-0008-9364-3289
capturing socioeconomic transformations and/or changes in patterns of land use and occupation occurring in intercensal periods, the dimensioning of service provision is compromised. Thus, unwanted impacts can be caused, such as the irregular disposal of solid waste, affecting the quality of life of the population. The use of alternative data sources, especially big data, has enormous potential to complement data from administrative records and surveys, in addition to filling information gaps generated by disconnection of data with reality as the period since the census reference date lengthens. In this article, procedures and results of the use of big data are presented, more precisely satellite images, as an information alternative capable of better reflecting collective demands for services such as collection of municipal solid waste, and therefore of better guiding the decision-making process at the municipal level.

**Keywords:** Big Data. SDG. Agenda 2030. Solid Waste.

**Resumo**

No Brasil, a coleta e o gerenciamento de resíduos sólidos urbanos são de responsabilidade dos governos municipais. Na maioria dos casos, isso é feito com base em informações, considerando os números da população para alocar recursos para essa finalidade. Sem dados capazes de captar transformações socioeconômicas e/ou mudanças nos padrões de uso e ocupação do solo ocorridas em períodos intercensitários, o dimensionamento da prestação de serviços fica comprometido. Dessa forma, impactos indesejáveis podem ser causados, como o descarte irregular de resíduos sólidos, afetando a qualidade de vida da população. A utilização de fontes de dados alternativas, especialmente de grandes volumes de dados, tem um enorme potencial para complementar os dados dos registros administrativos e dos inquéritos, além de preencher as lacunas de informação geradas pela desconexão dos dados com a realidade, à medida que o período decorrido desde a data de referência do recenseamento se prolonga. Neste artigo, são apresentados procedimentos e resultados da utilização de grandes volumes de dados, mais precisamente imagens de satélite, como uma alternativa de informação capaz de refletir melhor as exigências coletivas de serviços como a recolha de resíduos sólidos urbanos e, por conseguinte, de orientar melhor o processo decisório a nível municipal.

**Palavras-chave:** Big Data. ODS. Agenda 2030. Resíduos sólidos.
Introduction

The basic text of a bill of law that will guide the establishment of a National Plan for Smart Cities in Brazil is under debate in the Chamber of Deputies (lower house of Congress). This discussion is occurring in the wake of the elaboration of the Brazilian Smart Cities Letter (BRASIL, 2021), an initiative coordinated by the Ministry of Regional Development that involves, in addition to other institutions of the federal government, the support and expertise of international agents such as the German International Cooperation Agency (GIZ) and the International Telecommunication Union (ITU).

The federal government's agenda involving this issue has relevant dimensions, putting urban planning, sustainability and technology management at the interface.

Smart cities are cities committed to sustainable urban development and digital transformation, in their economic, environmental and socio-cultural aspects, which act in a planned, innovative, inclusive and networked way; promote digital literacy, collaborative governance and management; and use technologies to solve real problems, create opportunities, offer services efficiently, reduce inequalities, increase resilience and improve the quality of life of all people, ensuring the secure and responsible use of data and information and communication technologies (BRASIL, 2021, p. 28-29).

In parallel, in 2018 the National Bank for Economic and Social Development (BNDES) in partnership with the Ministry of Science, Technology and Innovation (MCTI), published the study "Internet of things: an action plan for Brazil". It contains another relevant dimension, related to big data, which presupposes the implementation of “global infrastructure for the information society, which enables provision of advanced services through the interconnection between things (physical and virtual), based on information and communication technologies (ICT). In the broad sense, it is not just about connecting things, but also about developing the power to process data today, called big data, making them 'intelligent'” (BNDES, 2018).

Big data are data from sources that can be generally described as: “high volume, velocity and variety of data that demand cost-effective, innovative forms of processing for enhanced insight and decision making.” Big data is characterized as datasets of increasing volume, velocity, and variety; big data is often largely unstructured, meaning that it has no pre-defined data model and/or does not fit well into conventional relational databases (Daas; Loo, 2013, p.1).

In order to assess the feasibility of the proposal, the bank financed some pilot initiatives in Brazilian cities, also considering the effects of this on optimization of the management of urban public goods and services.
Such efforts, carried out on a federal scale, fulfill a preparatory stage that precedes dissemination of the use of big data in connection with smart cities. To this end, in addition to the conceptual, regulatory, infrastructural and financial assumptions mentioned above, an important step concerns the diagnosis of Brazilian government maturity, especially at the municipal level, regarding the ability to access, process, interpret and transform the interpretive reading of big data into public policy actions. This is a concern that should be the object of future initiatives aimed at the technical training of public managers.

In statistical terms, measuring the degree of government maturity is important because the use of big data can improve the production of official statistics. In terms of urban management, the interplay of the smart cities framework, public information, pre-existing official data and big data should influence the planning process regarding the provision of certain services. This is valid both in terms of meeting the most immediate daily needs and demands of the population and in terms of macro-strategies aiming at future achievement of the Sustainable Development Goals (SDGs) - Agenda 2030, of the United Nations.

In this regard, SDG no. 11 – Sustainable Cities and Communities stands out, in particular its indicator 11.6.1: Proportion of municipal solid waste collected and managed in controlled facilities out of total municipal waste generated by cities (IBGE, 2023). The issue of municipal solid waste is also part of the framework of key performance indicators for smart cities, prepared by the International Telecommunication Union (ITU, 2021).

In the Brazilian case, the problem of municipal solid waste fits into the discussions of the National Solid Waste Policy (Law 12,305 of 2010) and the National Reverse Logistics Program (Presidential Decree 10,936 of 2022). In the current federative regime, it is up to the municipal governments to manage the collection of solid waste, either directly, indirectly by contracting private companies, or in consortiums with other municipalities.

Regardless of the modality, in most cases oversight is performed by control bodies based on information that considers the population of each city to allocate funds at the local level. These are in charge of understanding the spatial distribution of residents to offer the collection service, and if applicable, the treatment/disposal of waste.

Without having data capable of capturing socioeconomic transformations and/or changes in patterns of land use and occupation in intercensal periods, especially when there is delay in carrying out the census (as is the case of the 2020 Demographic Census), the sizing of solid waste collection services is compromised. Thus, unwanted impacts are caused, such as the irregular disposal of waste in inappropriate places, affecting the quality of life of the population.
population and moving the municipality away from the Sustainable Development Goals of the 2030 Agenda, of which Brazil is a signatory.

The use of alternative data sources, especially big data and administrative records, has enormous potential to complement surveys and census data, in addition to filling gaps due to the disconnection of data from reality, as the time since the census reference date passes. To support national statistical institutes in the use of big data, the United Nations created four regional hubs: Brazil, China, United Arab Emirates and Rwanda. The UN Regional Hub for Big Data in Brazil, which serves Latin America and the Caribbean, has been developing several actions to promote training, foster the interest of young statisticians in the use of big data, support the sharing of experiences and knowledge, strengthen ties, and promote cooperation in the region.

Among the initiatives supported by the Hub is the work presented here, which involves the development of procedures and the presentation of results from the use of big data, more precisely satellite images, combined with data from administrative records, to improve the estimation of the ODS Indicator 11.6.1: Proportion of municipal solid waste regularly collected and with adequate final discharge out of total municipal solid waste generated, in light of the reference of smart and sustainable cities. Therefore, we propose an informational and methodological alternative capable of better reflecting collective demands for services such as solid waste collection, and therefore of better guiding decision-making at the municipal level.

Given the intrinsic challenges related to the methodological limitations of this type of information, we chose the municipality of Maricá, in the state of Rio de Janeiro as the specific object of analysis. In this regard, we hope that this proving ground (Lipietz, 1988) will facilitate a future replication of the scope of the research now proposed in other locations, especially those located in the state of Rio de Janeiro.

**Identification of Irregular Disposal of Solid Waste Using Satellite Images**

The identification of areas with irregular disposal of solid waste through the inspection of free satellite images has been successfully carried out by different researchers for more than two decades. In the 1990s, Dafer et al. (1998) used satellite images and the ERDAS software

---

8 The municipality is the local administrative unit in Brazil. It is akin to a county, except with a single mayor and municipal council. Municipalities range from lightly populated rural ones with one or two small towns to heavily populated urban ones that are part of greater metropolitan regions. There are no unincorporated areas in Brazil.
to locate and monitor areas of irregular solid waste disposal in the city of Riyadh, Saudi Arabia. Four multispectral STOP satellite images at 1/100000 scale were used. The method used for the analysis after collecting the images was divided into four parts. The first corresponded to the restoration of images through geometric and radiometric correction. The second was to improve the quality of images by increasing sharpness and contrast. The third involved statistical analyses carried out using histograms based on variance-covariance matrices and on the frequency of the images, necessary for the selection, arrangement and analysis of the data. The fourth was the classification of images using the ERDAS software.

As a result, six landfills were identified in the region, five of them active and one inactive. It was also possible to visualize in the active landfills a different color in the soil, due to the decomposition of solid waste and the contamination of drinking water in the region.

In 2009, Yonezawa (2009) studied surface changes associated with irregular waste disposal using ALOS (Advanced Land Observing) images in the locality of Miyagi, Japan. The results indicated that a landfill could be identified, and temporal changes in the disposal site were clearly evident in the images. The author also analyzed images from the Quickbird satellite to assess the possibility of identifying residues of different sizes and materials. A junkyard measuring approximately 6x4 m was seen in the panoramic images captured in June 2003. The same junkyard, in an image obtained in September 2006, had expanded to approximately 10x4 m. One target was seen within a vegetated area, but another target of nearly the same size in an area with bare ground was difficult to see. A waste tire dump was employed to assess what could be identified from the Quickbird data. The author concluded that improvements in spatial resolution, shorter observation intervals and reductions in the cost of data could make satellite images very useful in monitoring the irregular disposal of solid waste.

Lanorte et al. (2017) proposed a method to identify, locate and quantify agricultural plastic waste using satellite images. The study area was the Apulia region in southern Italy, which is noted for intensive agriculture, so large amounts of agricultural plastic waste are generated from the disposal of films and nets for crop protection, irrigation pipes, pesticide containers and fertilizer bags. The authors used two methods to quantify plastic disposal in the ground: one based on the land use map available from the Apulia Region website and the other from the land use map created from the Landsat 8 satellite image. In both approaches, indices were calculated to quantify the generation of plastic waste for each type of cultivation and plastic application, in order to obtain the quantities of waste in the ground. The
A comparison of the results obtained with the two methods indicated a difference of 1.74% in the total amount of residues.

Akinina et al. (2017) used satellite images to identify areas of irregular waste disposal, three active and one inactive, and areas affected by contaminated water and pollution clouds caused by waste deposited in Riyadh, Saudi Arabia. Several images were examined until obtaining the best sharpness and contrast to facilitate the identification of areas. The classification of discard points was done automatically, using an algorithm from the Earth Resources Data Analysis System (ERDAS) Imagine software, which is a remote detection and photogrammetry application with raster graphics editor capabilities. The authors concluded that after refining the satellite images, it was possible to create an automatic image classifier for any aspect targeted for verification, suggesting that the use of this tool is interesting to optimize services such as the detection of solid waste in satellite images.

Gill et al. (2019) conducted a study on how to use the Land Surface Temperature gauge to determine landfill areas in the Al-Jaleeb region, Al-Farwanyah province, Kuwait. This could be observed because of the increase in soil temperature due to the decomposition of waste. The method used considered four steps: (i) data collection, using the Thematic Sensor of the Landsat Satellite Mapper, during the period between 1985 and 1994, divided into two categories: summer (July to August) and winter (January to March); (ii) image processing using the PCI Geomatica software, 2017 version; (iii) LST (Land Surface Temperature) computation to determine the areas with the highest temperature for identification of potential landfill areas; and (iv) accuracy assessment. As a result, greater accuracy was achieved in the summer months (72%) compared to the winter months (46%). Therefore, this detection method would only be recommended for use in the summer.

Vambol et al. (2019) studied the possibility of applying geoprocessing to reveal areas with irregular disposal of solid waste that could be potential raw materials for producing biofuels. The authors tested the proposed method at the Derhachi MSW landfill, located in the Kharkiv region of Ukraine. Based on the regulation of satellite image illumination and the pre-establishment of a certain texture of the images that would be the object of detection, they built an automatic classifier (based on size, location and shape of the detection object) of points with solid waste disposal. In addition, they applied statistical methods to detect which areas would be major generators of toxic gases. Due to the heterogeneity of the residuals, it was not possible to obtain high accuracy using a universal image classifier. However, the model developed based on the exclusion of low-risk disposal points allowed distinguishing disposal areas enriched with carbon-containing materials.
Skogsmo (2020) used free satellite imagery to identify irregular waste disposal in Kampala, Uganda, and Nairobi, Kenya. The author employed a version of Automatic Target Recognition (ATR), which is used in military applications, to detect and classify characteristics of areas with accumulations of solid waste. In addition, he used resources from Anaconda, Python, Jupyter notebook and the support vector machine (SVM) classification algorithm. The expectation was that the tests would demonstrate the viability of a classification model that can be improved with the increase in the number of sites and training data.

The experiments described do not exhaust the applications with the use of satellite images to identify and study areas with irregular disposal of solid waste, attesting to its viability.

**Municipal Solid Waste and the UN 2030 Agenda**

The 2030 Agenda for Sustainable Development is an “Action Plan for People, Planet and Prosperity [that] also seeks to strengthen universal peace in a context of greater freedom.” (United Nations, n.d.). To advance in the various dimensions of sustainable development, the Agenda establishes 17 Sustainable Development Goals, which address poverty and hunger eradication and sustainable agriculture, health and well-being, quality education, gender equality, safe water and sanitation, clean and affordable energy, decent work and economic growth, industry, innovation and infrastructure, reducing inequalities, sustainable cities and communities, responsible consumption and production, action against global climate change, life on water, life on land, peace, justice and effective institutions and partnerships and means of implementation.

Municipal solid waste is included in the SDG 11 objective - Sustainable Cities and Communities: to make cities and human settlements inclusive, safe, resilient and sustainable, and its production and destination can be monitored through indicator 11.6.1: Proportion of municipal solid waste collected and managed in controlled facilities out of total municipal waste generated by cities (UN Habitat, 2018). This indicator is defined by the following equation:

\[
I_{(11.6.1-UN)} = 100 \times \frac{v_1}{v},
\]

Where:

Revista Gestão e Secretariado (GeSec), São Paulo, SP, v. 14, n. 8, 2023, p. 14241-14261.
v1 is the volume of municipal solid waste regularly collected and properly disposed; and v is the volume of municipal solid waste generated.

Equation (1) can be rewritten as:

\[ I_{(11.6.1-\text{UN})} = 100 \times \frac{v_1}{v_1 + v_2 + v_3}, \]  

(2)

Where:

v2 is the volume of municipal solid waste collected regularly whose disposal is not adequate; and v3 is the volume of municipal solid waste not collected; with \( v_1 + v_2 + v_3 = v \).

Although apparently simple, calculating the indicator faces major difficulties in terms of data availability. In general, information on the volume of municipal solid waste regularly collected (\( v_1 + v_2 \)) can be obtained from local public agencies. However, separating the portion of this waste that has proper disposal (\( v_1 \)) is a little more complex, since it requires identifying whether the treatment facilities (composting, recycling, incineration, etc.) or disposal used have reached at least an intermediate level of control of environmental damage.

The evaluation of the suitability of disposal requires well-defined qualitative criteria and a technical team trained to carry out the monitoring. These conditions are not always met in municipalities. Thus, treatment or disposal facilities operated in an environmentally incorrect or unsatisfactory manner can be mistakenly classified as adequate destinations for solid waste, leading to overestimation of the indicator \( I_{(11.6.1-\text{UN})} \).

Obtaining the denominator of equation (2), municipal solid waste generated by the municipality, presents major obstacles. As previously mentioned, the volume of municipal solid waste collected regularly, regardless of destination, can be obtained from the local agencies. The volume of municipal solid waste not collected (\( v_3 \)) is, in general, unknown and needs to be estimated.

The United Nations (UN Habitat, 2018) proposes that this amount be estimated by multiplying the per capita volume of urban waste generated by the population that does not have access to collection service. In addition, it is suggested that, in the absence of information about this amount, a survey should be carried out in households, restaurants, hospitals, schools, etc. to calculate the daily generation of waste, through separation and weighing for 7 consecutive days, in at least two seasons of the year.
The implementation of this proposal faces two major challenges. The first is carrying out this survey at the municipal level. In addition to the operational complexity, a survey of this type, with the necessary frequency and geographic disaggregation, would be very costly in financial and personnel terms. The second challenge is to estimate the population without access to the service. Although this information is provided once every ten years by the Demographic Census in Brazil, accurate calculation of the indicator requires having current data every two years. Therefore, the search for alternative sources of data is crucial to guarantee the calculation of the indicator.

The SDGs in Brazil and the Indicator 11.6.1

In Brazil, indicator 11.6.1 - Proportion of municipal solid waste collected and managed in controlled facilities out of total municipal waste generated by cities is produced by the National Sanitation Secretariat, with data from the National Sanitation Information System (SNIS), self-declared annually by municipalities and basic sanitation service providers (ODSBRASIL, 2023). The calculation of the indicator, available annually for Brazil and broken down by its major regions, is defined by the following formula:

\[
I_{(11.6.1-BR)} = 100 \times \frac{v_{11} + v_{12}}{v_1 + v_2},
\]

Where:

- \(v_{11}\) is the volume of municipal solid waste regularly collected and disposed in landfills;
- \(v_{12}\) is the estimated mass of dry and organic recyclable solid waste recovered; such that \(v_{11} + v_{12} = v_1\).

In Brazil, the collected solid waste typically has various destinations, such as dumps, controlled landfills, sanitary landfills and/or recycling centers. According to the Ministry of the Environment (MMA, n.d.), a landfill is a place “with the proper environmental controls, such as the drainage of leachate and gases”, while “a dump has no control”, and a controlled landfill, as the name implies, “has a certain amount of control, but without a guarantee of environmental suitability”.

Law 12,305/2010, in Art. 3, numerals VII and VIII, defines the environmentally appropriate final destination as the “disposal of waste that includes reuse, recycling, composting, recovery and energy use or other destinations allowed by the competent bodies.”
Environmentally adequate final disposal, on the other hand, is equivalent to the “orderly distribution of waste in landfills, observing, specific operational standards in order to avoid damage or risks to public health and safety and to minimize adverse environmental impacts.” Thus, the choice of the numerator of indicator \( I_{(11.6.1-\text{BR})} \) is easily understood.

However, the choice of denominator raises some questions. When comparing the indicator proposed by the United Nations, \( I_{(11.6.1-\text{UN})} \), with the one calculated in Brazil, \( I_{(11.6.1-\text{BR})} \), it is clear that, in the latter, only the volume of solid waste collected is considered. Such a choice would only make sense if the volume of solid waste not collected was close to zero.

Although there are no data on the volume of garbage not collected, we can get an idea of this amount by observing the percentage of permanent private households in the urban area, whose garbage destination is different from that collected according to the household census. Although for the calculation of the indicator, other sources in addition to households must be considered, in general, households are more likely to be located in areas not covered by collection service.

According to data from the 2010 Demographic Census, the most recent at this writing (the 2020 Census was delayed), the percentage of permanent households in the urban area with waste collection was high, at 97.4%. However, this figure was different when considering municipalities. Although in 3,054 of the 5,565 Brazilian municipalities, the percentage was higher than in Brazil as a whole, in 82 municipalities fewer than 50% of permanent households in the urban area were served by collection and in 13 municipalities this was less than 20%. In the specific case of the municipality of Maricá, the percentage was lower than that of Brazil, 93.4%.

This information suggests that, for some municipalities, calculating the indicator using equation (3) would overestimate the result. Anyway, it is important to point out that at the moment, the Brazilian results for indicator 11.6.1 are available only for Brazil as a whole and the main regions.

Another point of reflection on the indicator is that although equation (3) involves the recovered mass of dry and recyclable organic solid waste, the three results made available to the public seem to say the opposite. They are:

- Proportion of mass for final disposal – landfill (%)
- Proportion of mass for final disposal – controlled landfill (%)
- Proportion of mass for final disposal – dump (%)

Revista Gestão e Secretariado (GeSec), São Paulo, SP, v. 14, n. 8, 2023, p. 14241-14261.
Since these three indicators add to 100%, it seems unlikely that recycling is being considered, either in the numerator or denominator of equation (3). Thus, we can rewrite the Brazilian indicator as:

\[ I_{(11.6.1-BR)} = 100 \times \frac{v_{11}}{v_{11} + v_2} \]  

Finally, despite the limitations, the information produced in Brazil for the SDG Indicator 11.6.1 is of great importance. From a statistical point of view, the estimates can be improved, seeking greater adherence to the indicator described by the United Nations (UN Habitat, 2018), which is municipal and whose denominator includes the portion of garbage not collected. In addition, it enables better calibrating the demand for collection/disposal of solid waste, allowing for faster revisions of the amount of intergovernmental resources passed on to municipalities, and thus ensuring effective implementation of the environmental agenda on the city scale.

### The Proposed Method

Our aim here is to improve the calculation of indicator \( I_{(11.6.1-BR)} \), presented in equation (4). For this pilot study, the municipality of Maricá was chosen. Because of its proximity to Niterói (fifth most populous municipality in the state) and especially to the capital city of Rio de Janeiro, this municipality has faced considerable challenges for solid waste management and conservation.

This is caused both by the fact that the population of Maricá fluctuates considerably between weekdays and weekends, and by the rapid restructuring of the pattern of land use and occupation, considering real estate investment projects that have expanded the urban area. From the point of view of governance, the city has responded with planning initiatives, with emphasis on the Municipal Sanitation Plan and the Municipal Plan for Integrated Solid Waste Management, carried out by municipally owned companies under the coordination of the municipal government.

The option for Maricá is justified by two main reasons. The first refers to the operational requirements related to the management of statistics that assume big data attributes, specifically the availability of open data related to public spaces. The second is the fact that the municipality in recent years has achieved remarkable socioeconomic dynamism.
and consequent changes in the pattern of territorial occupation, associated with the receipt of oil royalties (CUNHA et al., 2020) and public inclusion policies, such as the Mumbuca social currency (RODRIGUES; NEUMANN, 2021).

Although our final objective is to obtain an annual indicator, at this first moment we calculated the 11.6.1 indicator for a single month (November). Once the method is validated, it can be reproduced for the other months, providing an annual result. For this, we used two sources of data: information provided by the municipal government and data collection by inspection of satellite images.

More specifically, data referring to the volume of urban waste collected and properly disposed were provided by the municipal government and considered only disposal in landfills \(v_1\). As previously mentioned, this is a limitation, since the lack of data on recycling is common. According to the municipal government, 100% of the garbage collected in Maricá is deposited in landfills \(v_2 = 0\).

The novelty proposed here is the estimation of the volume of urban solid waste not collected \(v_3\), through inspection of satellite images, according to the following steps:

1. Team training for image inspection;
2. Identification of possible points of irregular disposal of municipal solid waste, that is, of uncollected waste, and demarcation of their areas with polygons;
3. Calculation of polygon areas;
4. Estimation of the height of polygons;
5. Calculation of the volume of waste not collected;
6. Selection of a sample of polygons to assess the quality of the work carried out;
7. Verification in the field of the selected polygons.

The estimated volume of uncollected waste \(\phi_3\) was calculated by adding the volumes of solid waste deposited in each polygon identified in the image inspection. The volume of waste in each polygon was obtained by multiplying the estimated area of the polygon by the estimated height of the waste accumulation:

\[
\phi_3 = \sum_{i=1}^{m} v_3 = \sum_{i=1}^{m} a_i c_i,
\]

Where:

\(m\) is the total number of polygons identified in the visual inspection of images;
\(i\) indicates the \(i\) - th polygon, where \(i=1, \ldots, m\);
\(a_i\) is the estimated area of the \(i\) - th polygon;
c_i is the estimated height of the accumulation of residues in the i-th polygon

Thus, the indicator for Maricá, I (11.6.1 - Maricá) can be written as:

\[ I_{11.6.1-Maricá} = 100 \times \frac{v_{11}}{v_{11} + v_2 + v_3} \]  \hspace{1cm} (6)

**Identification of Areas via Satellite**

The first step towards visual inspection of images was training with professionals in geoprocessing from the Brazilian Institute of Geography and Statistics (IBGE) to learn how to handle the Qgis software version 3.26 - Buenos Aires and to identify areas of municipal solid waste. The Gramacho landfill, in Duque de Caxias (RJ), and the slope and surroundings of Rua Tavares Bastos, in the city of Rio de Janeiro (RJ) – areas that have been proven to contain municipal solid waste – were used as training areas. In this and the following steps, images from Google Satellite, collected in November, were used.

Based on the acquired knowledge, layers were applied to the satellite images, identifying Brazil, the states, the municipality of Rio de Janeiro and the census tracts of Maricá. This process made it possible to isolate the area of interest, Maricá, dividing it into three areas of similar sizes (east, center and west) for the operationalization of the method.

Possible areas of irregular disposal were demarcated with polygons, calculating the respective areas using Qgis. Furthermore, the coordinates of a random point in the polygon were recorded so that they could be identified in the field. Such information, stored in Excel, enabled the selection of some of these polygons for field visits, based on their location.

**Indicator 11.6.1 Calculation with Administrative Records and Satellite Images**

According to data estimated for Maricá, in November 2022 about 8,000 tons of municipal solid waste was collected, all sent to a landfill, which, as seen earlier, is considered adequate disposal. Based on this information, we can calculate the indicator 11.6.1 by equation (4), reaching a value of 100%. This result shows that the entire mass of municipal solid waste generated in Maricá is collected and disposed of properly. However, satellite images and field verification belied this in many cases.

Adding the information obtained with the visual inspection of images, we recalculated the indicator by equation (6). To do so, it is first necessary to calculate the volume of irregular
disposal in square meters, according to equation (5) and then convert it into kilograms (kg) making it compatible with the information from the Maricá municipal government.

In all, 164 points of irregular disposal were identified in the urban areas of Maricá, totaling an area of 4,671 m². In addition, the average height of these areas was estimated at 0.5m, based on visits to some of the points identified in the image inspection. Using equation (5), the volume of solid waste discarded irregularly was estimated at 2,335 cubic meters (m³) or 700.6 tons, applying the method proposed by Zekkos et al. (2006) for converting waste volume to weight in tons.

Returning to equation (6), we have:

\[
\text{I}_{(11.6.1 - \text{Maricá})} = 100 \times \frac{v_{11}}{v_{11} + v_2 + v_3} = 100 \times \frac{8,000}{8,000 + 0 + 700.6} = 91.9
\] (7)

That is, we estimated that 91.9% of urban solid waste generated in Maricá is collected and sent to an appropriate location. The remaining volume is discarded irregularly. This result corroborates the statement that when using equation (4), the calculation of indicator 11.6.1, at least for Maricá, generates an overestimate.

**Validation of Visual Image Inspection**

To validate the visual identification of irregular disposal of solid waste, a sample of 15 points identified in the previous stage (8.9% of the total identified) was selected, throughout the Maricá, visited by the team responsible for the project during a day.

The sample was stratified by the three work areas of the visual inspection team, with selection proportional to size, that is, to the number of irregular waste disposal points identified in each of them. Table 1 shows the distribution of the points and the sample.

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Number of points identified</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>East area</td>
<td>89</td>
<td>8</td>
</tr>
<tr>
<td>Central area</td>
<td>56</td>
<td>5</td>
</tr>
<tr>
<td>West area</td>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>164</td>
<td>15</td>
</tr>
</tbody>
</table>

*Table 1: Number of irregular solid waste disposal points identified in the visual inspection and sample size selected for verification, by stratum - Maricá Urban Area - November 2022*

Source: Own elaboration.
In addition to the 15 points selected through the process described above, an irregular disposal point in the western area of Maricá was included in the sample with probability equal to 1, since it is the largest of all the areas identified in the visual inspection, covering 302.8 m².

The main objective of the field visits was to seek elements that would help assess the adequacy of the visual inspection of irregular waste disposal points (DIR). For this purpose, we created a scale that could be used during the field visits, which considered the following typology: municipal solid waste (MSW); civil construction waste (CCW); and green solid waste (GSW), which comes from pruning, weeding and/or clearing. This scale, with 8 categories, was defined as:

I. Points correctly identified in the visual inspection:
   1. Areas with MSW not collected;
   2. Areas with CCW or GSW that were contaminated and hence were considered to be MSW;
   3. Area with burned MSW;
   4. Area with MSW for recycling;
II. Areas with residues, but that did not fit the research objectives, that is:
   5. Areas with CCW;
   6. Private areas that could not be verified;
III. Waste-free areas
   7. Areas that did not contain waste, but that did contain traces of recent collection.
   8. Areas that did not contain waste or traces that during November had been repositories of irregular disposal.

It is important to point out that area types 1 to 4 must be included in the denominator of the proposed indicator, since they contain MSW generated and not collected by the municipal government. Therefore, without visual inspection, they would not be counted. The other areas should not be part of the calculation. Type 5 areas contain waste that is not part of the scope of the indicator. Type 6 areas are areas that cannot be scanned. Therefore, it was not possible to know if they really contained residues, and if so, if they were MSW. Type 7 areas underwent recent collection do their MSW volume had to be included in the total reported by the municipal government. Type 8 areas were those with mistakes in the visual inspection of images.

The field visit was carried out on December 21, 2022. The ENCE/IBGE team was accompanied by a technician from the Maricá Public Works Service (SOMAR). Due to
limited time (one day) and bad weather (constant rain throughout the day), only 9 of the 16 sample points were visited.

Of the 9 points, 6 were considered points correctly identified in the visual inspection. Of these, 3 were areas with MSW not collected and were classified with code 1; 1 contained CCW that was contaminated and given code 2; and 1 was an area with MSW for recycling and was classified with code 4. This point was inserted with a weight of 1 in the sample, since it had an area greater than 300 m$^2$. The last point, despite still containing a considerable amount of waste, was undergoing a burning process at the time of the visit. Thus, this point was classified with two different codes, 1 and 3.

Of the other 3 points visited, 1 was classified with code 6 for being a private area and therefore inaccessible to team; 1 with code 7, which appeared to have been recently cleaned; and 1 with code 8, as it was a flooded location that probably generated a water mirror effect in the satellite image, causing confusion in the identification of irregular waste disposal.

**Discussion**

The rise of big data refers to the context of informational intensification after the 1990s, when internet access became popular around the world. Since then, the evolution of the telecommunications infrastructure and computational skills and tools has consolidated those data as rich sources to produce statistics. This is the case with information obtained from mobile phone records, social media, websites, sensors, consumption meters, in addition to other sources. But what was most interesting in this study was satellite image repositories.

The use of big data has remarkable potential for supplementing public statistics. On the one hand, it opens up the possibility of comparing the traditional production of data – as is the case of the demographic censuses from which population estimates in Brazil are calculated – with more dynamic benchmarking for the dimensioning of phenomena. On the other hand, some challenges have emerged based on the unstructured nature of big data, such as the disruption of fundamental principles of statistics, the vulnerability to errors caused by selection biases and lack of representativeness, in addition to issues related to privacy, veracity and access. Although these are not issues exclusively related to big data, they are relevant to applications as described here.

That is why the use of big data requires the discussion of adequate references, especially with regard to methodological aspects, quality, use of technologies, accessibility, privacy regulations, management and storage of information and dissemination. Such
Concerns must be considered in order to attribute statistical validity to the data under analysis. Using them presupposes, therefore, a careful and judicious record of methodological procedures, whose purpose is to confer quality to the measurement and to the interpretations derived from the results found.

The foregoing is particularly relevant when dealing with use related to judging the quality of essential public services, such as collection of municipal solid waste through the examination of satellite images. For that, it has a consistent learning trajectory with previous experiences of using satellite images. The results reported here indicate a constant need for improvements in spatial resolution, shorter observation intervals, and above all reductions in the cost of obtaining images.

In the case of estimates related to municipal solid waste in the municipality of Maricá, although the increase in volume and the expansion of scope have been latent in recent years, it has been extremely difficult for the municipal government to measure and quantify them. This is due, above all, to the irregular way in which much of the disposal is carried out, but also to the mismatch between official collection coverage and the municipal socioeconomic dynamics.

This study showed that the proposed approach can be configured as an alternative form to improve measurement of public statistics. Indeed, the verification visit made by the research team demonstrated the validity of using satellite images, given the accuracy of the information referring to the location and quantification of the observed volumes. By estimating them more precisely, the municipal government can, for example, better substantiate claims related to the revision of criteria for intergovernmental revenue sharing from the federal and state levels for sanitation services and thus strengthen local governance dedicated to this end by expanding and optimizing collection.

Although there is potential to strengthen government capacity in this effort, the irregular disposal of urban waste is a permanent challenge. This waste can contain debris from civil construction, green waste, used furniture and other household items. These materials tend to be disposed irregularly in areas with little human traffic, such as the middle of woods and on the banks of rivers, where identification is difficult. Therefore, the use of satellite images can help locating such points to support cleanup efforts.

The study also opens perspectives for replicability in other contexts and government instances. As discussed, indicator 11.6.1 is an important metric to serve as a benchmark. Nevertheless, there are relevant technical difficulties, especially in emerging countries like Brazil, involving the availability of information related to the way the indicator is composed.
and governmental capacities – notably with local scale, as the case of Maricá illustrates. It is necessary to produce reliable estimates of the association between waste generation and economic and population dynamics. It is therefore urgent to close the gap imposed by the asymmetrical technological, financial and political conditions of public agents who wish to operate with big data.

In order to support political action aimed at improving the quality of public services, the use of big data must be considered as part of governance arrangements from the national to local contexts. In this respect, overcoming the technical and above all the financial limits faced by public managers will be facilitated.

It is also important to consider that the transformations observed in the spatiality of socioeconomic dynamics require that solid waste collection and disposal services be not only based on better estimates, such as those provided by satellite images, but also give rise to public-private governance arrangements and institutional frameworks better adapted to the demands of communities.

**Acknowledgement**

The authors thank everyone who contributed to this research. In particular, Lucy Mary Soares Souza, who helped us with the use of the QGIS tool, and the National Council for Scientific and Technological Development - CNPq for the scholarships granted to the authors Laira Ferreira and Fernanda Oliveira.

**References**


Submetido em: 17.07.2023
Aceito em: 17.08.2023