



## **Desafios e controvérsias no uso de big data em estatísticas oficiais: há uma luz no fim do túnel**

### **Challenges and controversies in the use of big data in official statistics: there's a light at the end of the tunnel**

DOI: 10.55905/revconv.16n.8-200

Recebimento dos originais: 24/07/2023

Aceitação para publicação: 24/08/2023

**Marcus André Alves Zimmermann Vieira**

Doutorando em População, Território e Estatísticas Públicas  
Instituição: Escola Nacional de Ciências Estatísticas (ENCE)

Endereço: Rio de Janeiro – RJ, Brasil

E-mail: marcusazimmermann@gmail.com

**Andrea Diniz da Silva**

Doutora em População, Território e Estatísticas Públicas  
Instituição: Escola Nacional de Ciências Estatísticas (ENCE)

Endereço: Rio de Janeiro – RJ, Brasil

E-mail: andrea.silva@ibge.gov.br

#### **RESUMO**

Nos últimos tempos, os institutos nacionais de estatísticas têm experimentado um aumento da demanda por estatísticas oficiais ao mesmo em que têm enfrentado cortes de orçamento. Neste cenário, o uso de big data como fonte de dados alternativa surge como uma opção mais ágil, eficiente e econômica para complementar fontes tradicionais de dados, como registros, pesquisas amostrais e censos. Contudo, o uso de big data também apresenta desafios, além de controvérsias, que não podem ser desconsideradas, em especial quando o interesse é produzir estatísticas oficiais. Dentre as principais questões quando se trata do uso de big data estão a representatividade, a preservação da privacidade, o acesso a tais dados, a individualidade e a autodeterminação do produtor primário, ou seja, o indivíduo que gerou o dado. O artigo pretende contribuir com esse debate, dialogando com a literatura nacional e internacional, compondo um cenário que permita os usuários de big data identificar cuidados necessários para que possam se beneficiar do uso desse tipo de dado de forma mais segura.

**Palavras-chave:** estatísticas oficiais, big data, desafios, controvérsias.

#### **ABSTRACT**

In recent times, national statistical offices have experienced an increased demand for official statistics while facing budget cuts. In this scenario, the use of big data as an alternative data source emerges as a more agile, efficient and economical option to complement traditional data sources, such as records, sample surveys and censuses. However, the use of big data also presents challenges, in addition to controversies, which cannot be ignored, especially when the interest is to produce official statistics. Among the main issues when it comes to the use of big data are representativeness, preservation of privacy, access to such data, individuality and self-



determination of the primary producer, that is, the individual who generated the data. The article intends to contribute to this debate, dialoguing with the national and international literature, composing a scenario that allows big data users to identify necessary care so that they can benefit from the use of this type of data in a safer way.

**Keywords:** official statistics, big data, challenges, controversies.

## 1 INTRODUÇÃO

Nos últimos tempos, os institutos nacionais de estatísticas têm enfrentado cortes de orçamento e, ao mesmo tempo, aumento da demanda por estatísticas oficiais. Neste cenário, o uso de big data como fonte de dados alternativa surge como uma opção mais ágil, eficiente e econômica. Logo, o uso de big data em estatísticas oficiais vem crescendo e apresentando considerável potencial para melhorar a sua quantidade e qualidade. Esse movimento motivou as Nações Unidas (ONU) a, por meio da Comissão Estatística, criar o Comitê de Especialistas da ONU em Big Data e Ciência de Dados para Estatísticas Oficiais (United Nations Committee of Experts on Big Data and Data Science for Official Statistics - UNCEBD, no original) em 2014 (UNSTATS, 2021). Dentre suas responsabilidades, o grupo é encarregado de fornecer visão estratégica, direção e coordenação de um programa global sobre o uso de big data para estatísticas oficiais. Esse grupo tem também a missão de promover o uso de fontes de big data para a produção de estatísticas oficiais, ao mesmo tempo que encontrar soluções para seus desafios, promover a capacitação e o compartilhamento de experiências. O grupo tem como guia fomentar a comunicação e a defesa do uso de big data para políticas públicas e para o monitoramento da Agenda 2030 para o Desenvolvimento Sustentável, além de construir a confiança do público no uso de big data para estatísticas oficiais (UNSTATS, 2021). Assim, o UNCEBD e a comunidade de estatísticos oficiais reconhecem a importância de abordar adequadamente as questões relacionadas a métodos, qualidade, tecnologia, acesso a dados, legislação, privacidade, gestão e finanças, além de fornecer análises de custo-benefício adequadas sobre o uso de big data (IBIDEM).

Além de organismos internacionais como a ONU, o Banco Mundial e o Fundo Monetário Internacional, institutos nacionais de estatística (INEs) têm investido em áreas dedicadas a descoberta de usos estatísticos para dados originalmente coletados com objetivos diversos, como os que estão disponíveis em fontes de big data. Biemer e Amaya (2020) relatam iniciativas bem



sucedidas nos INEs da Nova Zelândia, Holanda, Reino Unido e Canadá na produção de estatísticas oficiais a partir de big data. Esses INEs observaram que tais dados podem aumentar a relevância, reduzir a carga do respondente, aumentar a eficiência e produzir estatísticas mais oportunas e detalhadas (IBIDEM). Este é um passo importante no uso dessa fonte alternativa, assim como ocorreu nos países nórdicos, quando passaram a explorar registros administrativos tão bem que puderam utilizar a fonte, então considerada alternativa, para acompanhar a dinâmica sociodemográfica de suas populações.

Halderen et al (2021), compilaram exemplos de países que exploram big data, referida também como fontes de dados não tradicionais, para o monitoramento dos Objetivos de Desenvolvimento Sustentáveis (ODS). Tais dados incluem informações geoespaciais e dados gerados pelos cidadãos (este último tipo apenas pelas Filipinas). Ao todo, foram identificados 22 países explorando o uso de big data para monitorar a Agenda 2030 (Quadro 1). As experiências em curso mostram que o uso de big data na produção estatística é bastante promissor, quer seja para complementar ou para garantir a continuidade da produção estatística em cenários de restrição de recursos e de baixa cooperação dos respondentes.



Quadro 1 – Países que utilizam big data para o monitoramento dos ODS

País	ODS																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Alemanha						✓					✓				✓		
Austrália						✓								✓			
Áustria											✓				✓		
Canadá						✓					✓						
China	✓	✓	✓			✓			✓		✓			✓	✓		
Colômbia									✓		✓						
Espanha															✓		
Filipinas	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	
Finlândia											✓				✓		
França											✓				✓		
Holanda						✓											
Índia											✓						
Irlanda											✓						
Itália											✓				✓		
Japão									✓						✓		
México						✓					✓				✓		
Nova Zelândia															✓		
Portugal											✓						
Suécia											✓						
Suíça											✓						
Turquia															✓		
Ucrânia	✓														✓		

Fonte: Adaptado de Halderen et. al. (2021)

## 2 FUNÇÃO PRECÍPUA E REPRESENTATIVIDADE

Apesar dos benefícios do uso de big data, seu uso é controverso, pois apresenta desafios, tendo em vista que normalmente, suas fontes, diferentemente de censos e pesquisas amostrais, não são primariamente estruturadas para fornecer estatísticas, assim como acontece com os registros administrativos (BIEMER e AMAYA, 2020). Ou seja, a “derivação de estatísticas é um propósito secundário” (MACFEELY, 2019). Muitas vezes esses dados são não estruturados, pois são criados da “exaustão” do sistema sem nenhum propósito específico que não seja a preservação de dados (BIEMER e AMAYA, 2020) ou mesmo com finalidade administrativa. Não obstante, a comunidade estatística percebeu o potencial dessas bases enquanto fontes de dados. Tam e Kim (2018) lembram que, por meio de pesquisas, as instituições possuem controle da coleta de dados, incluindo conteúdo (o quê), forma (como) e período (quando), os quais são produzidos com o objetivo principal de produzir estatísticas. Diferentemente, big data normalmente é composto de dados gerados automaticamente, o que torna as estatísticas





produzidas com base em tais fontes um subproduto do mesmo, não sendo sua principal orientação (MACFEELY, 2019).

Dependendo da população alvo, big data também pode não apresentar a melhor representatividade da mesma, o que, por sua vez, pode comprometer as estimativas das variáveis de interesse. Isso ocorre, por exemplo, porque algumas fontes de big data podem não incluir as pessoas “não digitais”. Pensando exclusivamente nas transações financeiras, é sabido que uma parcela da população não tem conta bancária e executa suas operações financeiras apenas em papel moeda. Sendo assim, as transações mapeadas pelo sistema financeiro, por meio de controles de relatórios dos bancos comerciais, não incluem aquelas realizadas por essas pessoas. Assim, não seria suficiente, usar tal base para fazer inferências sobre toda a população brasileira, devido a não representatividade desse grupo. Nesse sentido, necessitar-se-ia ajustes para corrigir o viés de seleção<sup>1</sup>, para garantir a generalização. Isso vale também para os dados apurados considerando as pessoas conectadas à internet, tendo em vista que uma parcela da população continua *off-line*. Vale pontuar que a UIT (União Internacional de Telecomunicações), estima que, em 2021, cerca de cinco bilhões de pessoas usavam internet, o que representa 63% da população mundial (UIT, 2022), enquanto o Cetic.br (Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação) aponta que o Brasil, em 2021, possuía cerca de 148 milhões de usuários, o que representa 81% da população brasileira com 10 anos ou mais de idade (CETIC.BR, 2022).

### 3 PRIVACIDADE, CONFIDENCIALIDADE, INDIVIDUALIDADE E AUTODETERMINAÇÃO

Privacidade e confidencialidade são preocupações que se estendem também ao uso de big data. Para Kim e Tam (2018), o primeiro é referente ao direito de controle de informações pessoais e da própria vida pessoal livre de intrusões, enquanto o segundo é uma obrigação do controlador das informações em mantê-las secretas, evitando a sua divulgação. No Brasil, a privacidade é tratada na Lei 13.709/2018, a Lei Geral de Proteção de Dados (LGPD), que “dispõe

---

<sup>1</sup> O viés de seleção ocorre quando há uma falha na seleção de unidades ou objetos de estudo. Esse tipo de situação pode ocorrer quando se favorece algum grupo, por exemplo, e pode ocorrer devido a algum método de coleta que privilegie certo grupo. Com isso, pode haver falhas de representatividade, afinal a amostra selecionada pode não ser proporcional à população, incorrendo em uma falha de cobertura, em que a amostra selecionada difere sistematicamente da população. Logo, qualquer inferência sobre a mesma para uma população mais geral estaria distorcida. Vale pontuar também que, como o viés de seleção é um erro não amostral, o mesmo não pode ser reduzido aumentando o tamanho da amostra, como já fora demonstrado por Tam e Kim (2018).



sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural” (BRASIL, 2018). A lei complementa o Marco Civil da Internet (Lei 12.965/2014), que “estabelece princípios, garantias, direitos e deveres para o uso da internet no Brasil e determina as diretrizes para atuação da União, dos Estados, do Distrito Federal e dos Municípios em relação à matéria.” (BRASIL, 2014). Em seu art. 7, inciso X, a lei prevê o direito do cidadão brasileiro à exclusão definitiva dos dados pessoais que tiver fornecido em determinada aplicação de internet, a seu requerimento, ao término da relação entre as partes, excluindo os casos em que é obrigatória a guarda de registros, previstos nesta lei. Já a confidencialidade é tratada nos Princípios Fundamentais das Estatísticas Oficiais da ONU, especificamente pelo princípio 6, que preconiza que “os dados individuais recolhidos pelos organismos estatísticos para fins de compilação estatística, quer se refiram a pessoas físicas ou jurídicas, são estritamente confidenciais e utilizados exclusivamente para fins estatísticos” (NAÇÕES UNIDAS, 2014). O mesmo também está presente no Código de Boas Práticas das Estatísticas do IBGE, revisado em 2021. A confidencialidade assegurada pelos INEs é necessária para garantir a confiança nessas instituições e pode garantir a credibilidade na relação entre os respondentes e o sistema estatístico nacional.

A preservação da privacidade representa um desafio adicional quando se trata do uso de big data. Pereira e Cachapuz (2018) alertam para a dificuldade de prever de antemão quais conexões serão realizadas, ou para que finalidade esses dados serão utilizados. Isso porque, muitas das vezes, coleta-se o dado sem ter dimensão do seu potencial. Hoje várias corporações coletam informações pessoais sem informar com exatidão como serão utilizadas. Podem inclusive coletar informações com um objetivo e usá-lo para fins diversos, até mesmo comerciais, já que os dados podem ser reprocessados. Isso, mesmo considerando um cenário de boa-fé entre as partes. É o que MacFeely (2019) chama de “reciclagem” do dado para reaproveitá-lo. Com isso, as estratégias previstas em lei para proteger os dados, como consentimento individual, opção de exclusão e anonimização dos dados, tornam-se frágeis. As autoras observam que, diante da dificuldade de precaver o mau uso dos dados, resta focar na punição e responsabilização por uso indevido ou abusivo dos dados, apesar de isso não garantir a privacidade.



MacFeely (2019) ainda ressalta o desafio adicional quando se trata de dados cujo escopo é supranacional ou global, como ocorre com muitos dados digitais, como imagens de satélite e sensores de navios. Teoricamente nesses casos, seria mais eficiente contar com um modelo centralizado na gestão desses dados em vez do modelo atual, em que cada instituto nacional produz e utiliza tais dados. Um modelo centralizado possibilitaria ganhos, inclusive, no aspecto de comparabilidade entre os países. Contudo, essa colaboração internacional pode representar um risco a soberania dos países, pois os dados estariam fora da proteção dos sistemas normativos nacionais. Nesse sentido, dever-se-ia encontrar o equilíbrio entre a maior eficiência na gestão de dados supranacionais e a garantia da soberania dos países. Esse é um tema em discussão, cuja relevância aumentou nos últimos anos, com o crescimento do uso de big data.

Pereira e Cachapuz (2018) alertam também que são comuns a colisão de princípios contrários, ou seja, quando ambos são considerados fundamentais e, portanto, não poderiam ser desrespeitados, mas se contrapõem. Nesses casos, o caminho legal apontado pelas autoras é o da ponderação, que necessita da avaliação segundo três critérios: adequação, necessidade e proporcionalidade. Um exemplo usado pelas mesmas é o caso de contraposição do direito à intimidade versus o interesse público em relação à segurança do Estado. Isso porque, todos têm direito a terem sua intimidade preservada, mas se isso representar uma ameaça terrorista, por exemplo, a garantia da segurança pública é priorizada em detrimento ao direito à intimidade. Neste aspecto, Schroeder (2018) também concorda que certas liberdades podem ser restringidas em prol de outras causas, como a segurança. Fica, então, o desafio de lidar com o conflito de interesses quando se trata de gestão e prevenção de riscos, enfrentamento de crises sanitárias, gestão da produção e distribuição de alimentos, segurança pública, entre outras aplicações do uso de big data já experimentadas com sucesso.

Pereira e Cachapuz (2018) advertem para o fato de que as informações coletadas na forma de big data podem ser utilizadas por detentores privados para influenciar opiniões ou mesmo estimular determinados padrões de consumo, o que pode ferir não só o direito à privacidade, mas também a própria liberdade individual. Quem nunca recebeu sugestões ou recomendações de compras ou conteúdos? Isso é fruto de um algoritmo que analisa padrões e estipula seus gostos. Contudo, nada impede que isso evolua para moldar essas preferências, o que permitiria influenciar opiniões e compras, intervindo no comportamento humano (SCHROEDER, 2018). As autoras ainda apontam um mecanismo muito frequente utilizado em diversas aplicações, que





praticamente obrigam a aceitação dos termos de uso, que, na maioria das vezes, autoriza as empresas a utilizar dados pessoais. Assim, o usuário troca a sua privacidade por determinados benefícios como o uso gratuito de aplicativos ou conteúdos gerados pela empresa. Schroeder (2018) ainda reforça tal ideia ao criticar um estudo sobre contágio emocional realizado pelo Facebook, que evidencia a capacidade de um algoritmo em manipular o humor de usuários de redes sociais. Vale ressaltar que tal estudo ainda foi antiético ao não contar com o consentimento esclarecido dos analisados.

Tais práticas também entram em conflito com outro conceito, o de autodeterminação informativa, que consiste no “pleno controle sobre as informações normativas de cada um”, ou seja, o controle do indivíduo de seus próprios dados pessoais (PEREIRA e CACHAPUZ, 2018, p. 1068). A autodeterminação informativa tem a função de permitir o livre desenvolvimento do comportamento do indivíduo. Em um cenário, onde cada vez mais somos induzidos a nos sujeitar a termos de uso ou mesmo estimulados a expor conteúdos íntimos nas redes sociais, garantir o direito à intimidade, assim como da privacidade fica mais difícil. Schroeder (2018) ainda complementa que o uso comercial de big data põe em xeque nossas noções de individualidade e autodeterminação. Para ele, se o Facebook almejar prever a personalidade de alguém ou o seu comportamento, isso pode não ser disruptivo, mas pode prejudicar seu senso de individualidade em um nível pessoal. Adicionam-se ainda manipulações que ferem decisões pessoais ou coletivas, minimizando a importância do livre arbítrio. Ainda que essa influência também possa ter aplicações positivas, como mudar hábitos de consumo de energia em um cenário de graves mudanças climáticas, por exemplo (SCHROEDER, 2018), é necessário lidar com os efeitos negativos.

#### **4 ACESSO AOS DADOS E OUTRAS CONTROVÉRSIAS**

MacFeely (2019), Pereira e Cachapuz (2018), Kim e Tam (2018) e Schroeder (2018) alertam para o fato de que parte significativa dos dados classificados como big data é de propriedade privada e sujeito a custos, o que dificulta seu uso para estatísticas públicas, apesar do evidente interesse público, limitando assim o seu potencial. Em certas bases, há, inclusive, legislação que impede o acesso. Também, muitas bases de big data representam capital estratégico de corporações comerciais que não cederiam gratuitamente esses dados. Por essa razão, substituir os dados tradicionais por big data pode comprometer a produção de estatísticas





oficiais, por gerar uma maior dependência de fontes externas e fazer os INEs perderem o controle de seu sistema de produção. A substituição poderia expor os sistemas estatísticos nacionais a choques exógenos, como problemas com as empresas detentoras desses dados, ou mesmo um aumento de preço para o compartilhamento dos dados, além da perda de controle quanto à amostragem e ao processamento dos dados (MACFEELY, 2019).

MacFeely (2019) também aponta que algumas fontes de big data apresentam volatilidade e instabilidade, o que pode comprometer a integridade e a continuidade de fornecimento do dado. Isso deve ser levado em conta na manutenção de séries históricas. Afinal, como a tecnologia está em constante evolução, os dados coletados podem ser modificados a qualquer momento. O autor ainda se preocupa com a reputação das instituições que usam big data para geração de estatísticas, pois sua confiança e reputação pública podem ser questionadas no caso de infração de leis ou mesmo ultrapassar limites culturais e morais de uma sociedade. O autor alerta que todo cuidado é pouco para garantir que o nome dessas instituições seja preservado, pois qualquer quebra de confiança, pode prejudicar a sua reputação, principalmente em regiões onde não haja ainda leis claras sobre o assunto.

Por fim, mas não menos importante, vale mencionar que nem tudo capturado em fontes de big data, representa uma pessoa ou instituição real. MacFeely (2019) apresenta uma série de estatísticas que evidenciam traços de contas em mídias sociais classificadas como falsas ou até robôs, além das duplicações. Muitas das interações apuradas são realizadas por robôs, como, por exemplo, o que é feito por meio de web scraping. Neste caso, para essa informação poder ser utilizada para uma análise mais adequada, esta parte do tráfego extra deveria ser desconsiderada.

## 5 DISCUSSÃO

O uso de big data em estatísticas oficiais não é direto e livre de cuidados. Há que se atentar que majoritariamente essas bases não foram criadas para geração de estatísticas públicas, então podem apresentar problemas de representatividade, que podem influenciar diretamente na geração de estimativas devido ao viés de seleção. Outra preocupação é quanto à garantia à manutenção da privacidade e confidencialidade dos dados, ainda frágil mesmo com leis e princípios de boas práticas. Para tanto, é necessário observar a dificuldade de previsão quanto ao uso dos dados coletados sob consentimento compulsório, tendo em vista que os mesmos podem ser reprocessados para outros fins, necessitando de estratégias fortes de punição para reprimir



o mau uso dos dados. A dependência de fontes externas aos produtores de estatísticas oficiais também pode comprometer o controle da produção em razão de eventuais problemas nas empresas detentoras dos dados, além de possíveis flutuações nos custos de sua transferência. Adicionam-se questões legais que podem envolver cláusula protetivas ou restritivas que afetem o compartilhamento de dados, já implantadas em vários países. Outras questões, como volatilidade e instabilidade de algumas fontes, assim como a presença de robôs, contas falsas e duplicadas na geração de tal tipo de dado, adicionam desafios para o uso de big data.

A busca de estratégias para o enfrentamento dos desafios (não exclusivos) de big data está em curso. Biemer e Amaya (2020) propõem a combinação de big data com pesquisas amostrais tradicionais para reduzir o viés da primeira e resolver alguns dos problemas de qualidade de dados em dados de pesquisa. Kim e Wang (2018) tratam o viés de seleção em big data como um problema de dados ausentes e para resolver isso, propõem dois métodos. O primeiro método utiliza uma versão de amostragem inversa incorporando informações auxiliares de fontes externas e o segundo usa a ideia de integração de dados, combinando big data com uma amostra probabilística, baseado em método de ponderação. A partir de dois estudos de simulação, os autores mostram que os métodos propostos são imparciais e têm melhores taxas de cobertura do que o uso de apenas uma das fontes de dados, além de serem de fácil implementação.

Quanto ao tratamento da confidencialidade e privacidade, a força-tarefa Tecnologias para a Melhoria da Privacidade<sup>2</sup>, criada em 2018 para assessorar o UNCEBD na questão, lançou um manual de técnicas de preservação da privacidade. Nesta documentação, foram descritas motivações para abordagens de preservação da privacidade para a análise estatística de dados confidenciais, exemplos de casos de uso em que tais métodos podem ser utilizados e capacidades técnicas importantes para garantir a preservação da privacidade enquanto permite a análise de dados confidenciais. Dentre as técnicas descritas que suportam a proteção e o compartilhamento de informações confidenciais, destacam-se: Computação Multipartidária Segura, Criptografia Homomórfica, Ambientes de Execução Confiáveis, Privacidade Diferencial e Provas de Conhecimento Zero (ONU, 2023).

Outra força-tarefa que assessora o UNCEBD quanto ao uso de big data, discute a facilitação do acesso global a dados de propriedade privada para estatísticas oficiais, já que, como já foi debatido, o acesso a dados privados mostra-se como um desafiador problema. Criada em

---

<sup>2</sup> Tradução livre de Privacy-Enhancing Technologies Task Team (PETTT)



2021, a mesma possui três funções principais: (i) negociar acesso global a fontes de big data privados sob acordos que prevejam o seu uso estritamente para fins estatísticos, especialmente para avançar na implementação da Agenda 2030 para o Desenvolvimento Sustentável; (ii) demonstrar a relevância dessas fontes de dados para fins estatísticos para a comunidade estatística oficial; e (iii) aconselhar sobre arranjos institucionais após a conclusão de experimentos e testes bem-sucedidos (UNSTATS, 2023).

Logo, o uso de big data para melhorar a produção de estatísticas requer um gerenciamento de expectativas. Ainda há muito trabalho a ser feito para que o potencial das diversas fontes de big data possa ser utilizado integralmente na produção de estatísticas oficiais. Questões como representatividade, confidencialidade e acesso aos dados estão sendo discutidas pela comunidade estatística internacional e diferentes soluções vêm sendo experimentadas. Contudo, é necessário um tempo para alcançar maturidade e consolidar métodos e técnicas. O presente artigo não esgota os desafios e as controvérsias em torno do tema, assim como os esforços que vêm sendo feitos, porém dá visibilidade ao potencial de big data para ser usado em conjunto ou como substituto de fontes de dados tradicionais para aumentar, melhorar e complementar as estatísticas oficiais, tornando o sistema produtor mais ágil, eficiente e econômico, isso sem contar as possibilidades de desagregação e pareamento de dados.

### **AGRADECIMENTOS**

O autor principal agradece ao IBGE e a Capes pelas bolsas de estudo concedidas, o que possibilitou o presente trabalho de pesquisa.





## REFERÊNCIAS

BIEMER, P. P.; AMAYA, A. (2020). Total Error Frameworks for Found Data. 10.1002/9781118976357.ch4.

BRASIL. (2014). Lei nº 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2014/lei/l12965.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm). Acesso em: 24/11/2021.

BRASIL. (2018). Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/l13709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm). Acesso em: 24/11/2021.

CETIC.BR. (2022). TIC Domicílios 2021. Lançamento dos resultados. Disponível em: [https://cetic.br/media/analises/tic\\_domicilios\\_2021\\_coletiva\\_imprensa.pdf](https://cetic.br/media/analises/tic_domicilios_2021_coletiva_imprensa.pdf). Acesso em 26/10/2022

HALDEREN, G. V.; BERNAL, I.; SEJERSEN, T.; JANSEN, R.; PLOUG, N.; TRUSZCZYNSKI, M. (2021). Big Data for the SDGs. Country examples in compiling SDGbig data para a produção estatística indicators using non-traditional data sources. Working Paper Series. ESCAP Statistics Division. SD/WP/12. Disponível em: [https://www.unescap.org/sites/default/d8files/knowledge-products/SD\\_Working\\_Paper\\_no12\\_Jan2021\\_Big\\_data\\_for\\_SDG\\_indicators.pdf](https://www.unescap.org/sites/default/d8files/knowledge-products/SD_Working_Paper_no12_Jan2021_Big_data_for_SDG_indicators.pdf). Acesso em 24/07/2023.

IBGE. (2021). Código de Boas Práticas das Estatísticas do IBGE. 2ª edição. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101744.pdf>. Acesso em: 24/11/2021.

KIM, J. K.; WANG, Z. (2018). Sampling Techniques for Big Data Analysis. International Statistical Review. 87. 10.1111/insr.12290. Disponível em: [https://www.researchgate.net/publication/327351727\\_Sampling\\_Techniques\\_for\\_Big\\_Data\\_Analysis](https://www.researchgate.net/publication/327351727_Sampling_Techniques_for_Big_Data_Analysis). Acesso em: 28/07/2023.

MACFEELY, S. (2019). The Big (data) Bang: Opportunities and Challenges for Compiling SDG Indicators. Global Policy. 10. 121-133. 10.1111/1758-5899.12595. Janeiro, 2019.

NAÇÕES UNIDAS. (2014). Fundamental Principles of Official Statistics. Resolution adopted by the General Assembly on 29 January 2014. 73rd plenary meeting. Disponível em: <https://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf>. Acesso em: 07/08/2023.

ONU. (2023). UN Privacy Preserving Techniques Handbook. Disponível em: <https://unstats.un.org/bigdata/task-teams/privacy/UN%20Handbook%20for%20Privacy-Preserving%20Techniques.pdf>. Acesso em: 09/08/2023.



PEREIRA, M. V.; CACHAPUZ, M. C. (2018). Big data e o conflito entre a utilização dos dados e a proteção à intimidade e a vida privada. RJLB, Ano 4, nº 1, p. 1067-1085. Disponível em: [https://www.cidp.pt/revistas/rjlb/2018/1/2018\\_01\\_1067\\_1085.pdf](https://www.cidp.pt/revistas/rjlb/2018/1/2018_01_1067_1085.pdf). Acesso em: 12/03/2021.

SCHROEDER, R. (2018). Big data: moldando o conhecimento, moldando a vida cotidiana. Matrizes V.12, Nº 2, maio/ago, p. 135-163. São Paulo, Brasil. DOI: <http://dx.doi.org/10.11606/issn.1982-8160.v12i2p135-163>.

TAM, S. M.; KIM, J. K. (2018). Big Data ethics and selection-bias: An official statistician's perspective. Statistical Journal of the IAOS. 34. 1-12. 10.3233/SJI-170395. Disponível em: <https://content.iospress.com/download/statistical-journal-of-the-iaos/sji170395?id=statistical-journal-of-the-iaos%2Fsji170395>. Acesso em: 19/10/2021.

UIT. (2022). Statistics. Disponível em: <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>. Acesso em: 31/03/2022.

UNSD. (2014). Fundamental Principles of Official Statistics. Disponível em: <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>. Acesso em: 24/11/2021.

UNSTATS. (2021). UN Committee of Experts on Big Data and Data Science for Official Statistics. Mandate and Terms of Reference of the UNCEBD. Disponível em: <https://unstats.un.org/bigdata/about/mandate.cshtml>. Acesso em: 08/12/2021.

UNSTATS. (2023). Global Facilitation of Access to Privately-held Data for Official Statistics. Joint Task Team. UN Committee of Experts on Big Data and Data Science for Official Statistics. UN Network of Economic Statisticians. Disponível em: <https://unstats.un.org/bigdata/task-teams/data-access/index.cshtml>. Acesso em: 09/08/2023.