Reciis – Revista Eletrônica de Comunicação, Informação & Inovação em Saúde, Rio de Janeiro, v. 16, n. 3, p. 742-745, jul.-set. 2022 [www.reciis.icict.fiocruz.br] e-ISSN 1981-6278

https://www.doi.org/10.29397/reciis.v16i3.3471

RESENHAS

A pesquisa científica na era do *Big data*: cinco maneiras que mostram como o *Big data* prejudica a ciência, e como podemos salvá-la

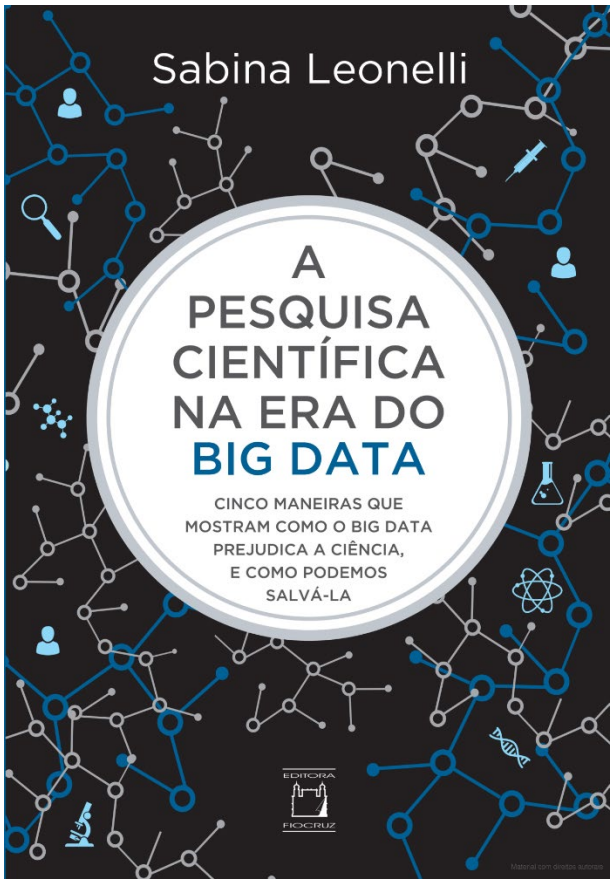
The scientific research in the age of Big data: five ways that show how the Big data harms the science, and how we can save it

La investigación científica en la era del *Big data*: cinco maneras que muestran como el *Big data* perjudica la ciencia, y como la salvar

Raphael de Freitas Saldanha[1,a]
raphael.saldanha@icict.fiocruz.br | https://orcid.org/0000-0003-0652-8466

[1] Fundação Oswaldo Cruz, Instituto de Comunicação e Informação Científica e Tecnológica em Saúde. Rio de Janeiro, RJ, Brasil.

[a] Doutorado em Informação e Comunicação em Saúde pela Fundação Oswaldo Cruz.



RESUMO

O livro A pesquisa científica na era do *Big data*: cinco maneiras que mostram como o *Big data* prejudica a ciência, e como podemos salvá-la, de Sabina Leonelli, publicado pela Editora Fiocruz em 2022, explora em seus capítulos as definições do termo *Big data* e os seus impactos negativos na pesquisa científica. Em seguida, a autora revela uma nova abordagem epistemológica para o *Big data* e, por fim, apresenta um conjunto de propostas para a pesquisa científica. A revisão e atualização de definições, tanto quanto as importantes reflexões e os questionamentos por um uso consciente do *Big data* na pesquisa científica fazem com que a obra adicione importantes contribuições à biblioteca do pesquisador de informação e comunicação em saúde.

Palavras-chave: *Big data*; Ciência de dados; Saúde pública; Epidemiologia; Sistemas de Informação em Saúde.

ABSTRACT

The book titled A pesquisa científica na era do Big Data: cinco maneiras que mostram como o Big Data prejudica a ciência, e como podemos salvá-la [The scientific research in the age of Big Data: five ways that show how the Big Data harms the science, and how we can save it], by Sabina Leonelli, published in 2002, by Editora Fiocruz, explores in its chapters the definitions of Big Data and its negative impacts on scientific research. Then, the author reveals a new epistemological approach to Big data and finally she presents a set of proposals for developing a good scientific research. The literature review and updating of definitions as well as the important reflections and questions for a conscious use of Big data in scientific research make the work an important contribution to the researcher's library of the information and communication about health.

Keywords: Big data; Data science; Public health; Epidemiology; Health Information Systems.

RESUMEN

El libro denominado A pesquisa científica na era do *Big data*: cinco maneiras que mostram como o *Big data* prejudica a ciência, e como podemos salvá-la [La investigación científica en la era del *Big data*: cinco maneras que muestran como el *Big data* perjudica la ciencia, y como la salvar], de Sabina Leonelli, publicado en 2002, por la Editora Fiocruz, explora em sus capítulos las definiciones de *Big data* y sus impactos negativos en la investigación científica. A continuación, la autora revela un nuevo enfoque epistemológico del *Big data* y, al fin y al cabo, presenta un conjunto de propuestas para desarrollar una investigación científica de cualidad. La revisión de literatura y la actualización de las definiciones, así como las importantes reflexiones y discusiones para un uso consciente del *Big data* en la investigación científica, hacen de la obra un aporte importante a la biblioteca del investigador de la información y la comunicación acerca de la salud.

Palabras clave: *Big data*; Ciencia de datos; Salud pública; Epidemiología; Sistemas de Información en Salud.

INFORMAÇÕES DO ARTIGO

Obra resenhada: LEONELLI, Sabina. **A pesquisa científica na era do *Big data*:** cinco maneiras que mostram como o *Big data* prejudica a ciência, e como podemos salvá-la. Tradução de: Carla Cristina Munhoz Xavier. Rio de Janeiro: Editora Fiocruz, 2022.

Contribuição dos autores: o autor é responsável por todo o texto.

Declaração de conflito de interesses: não há.

Fontes de financiamento: não houve.

Considerações éticas: não há.

Agradecimentos/Contribuições adicionais: não há.

Histórico do artigo: submetido: 26 set. 2022 | aceito: 26 set. 2022 | publicado: 30 set. 2022.

Apresentação anterior: não houve.

Apresenta-se aqui uma resenha do livro A pesquisa científica na era do *Big data*: cinco maneiras que mostram como o *Big data* prejudica a ciência, e como podemos salvá-la, da autora italiana Sabina Leonelli e tradzido por Carla Cristina Munhoz Xavier, publicado em 2022 pela Editora Fiocruz.

O título pode causar certa estranheza em um primeiro momento. A tese enfática já no título de que o *Big data* prejudica a ciência e, logo em seguida, oferecer a solução para este problema pode parecer simplista. Contudo, a obra muito bem documenta a tese proposta, além de propor possíveis encaminhamentos e soluções para os problemas pontuados.

A obra se divide em quatro capítulos básicos que exploram as definições do termo *Big data* e os seus impactos negativos na pesquisa científica. O texto traz uma nova abordagem epistemológica para o termo e, por fim, apresenta um conjunto de propostas para a ciência que se apoia no *Big data*.

A relevância dessa obra de Leonelli para a informação e comunicação em saúde no Brasil se dá pela proximidade temporal da crescente adoção de tecnologias e abordagens de *Big data* na área da saúde. Na variedade das questões que Leonelli aborda nos capítulos não é difícil encontrar contrapontos e exemplos de situações existentes na saúde pública brasileira.

Para Leonelli, *Big data* pode ser definido como: "[...] dados de diferentes tipos e origens que se relacionam, muitas vezes em formato digital e de formas que se prestam ao aprendizado de máquina, de modo a produzir novos procedimentos de análise e conhecimento." (p. 25)

Em sua definição, Leonelli incorpora – ainda que indiretamente – a ideia de que *Big data* é uma caracterização dos dados utilizados pela ciência em aplicações metodológicas mais recentes, em um universo conceitual de uma ciência de dados.

Ao discutir sobre os conceitos de "ciência aberta" (p. 26), Leonelli destaca a revolução corrente na comunicação dos resultados de pesquisa, tema de crescente discussão na saúde pública (SALES, 2021).

A autora também reflete sobre a evolução histórica da utilização de dados na ciência, conceituando *Big data* como o momento em que o dado se torna "componente fundamental e resultado da pesquisa científica" (p. 30).

A utilização de grandes bases de dados nos estudos da saúde pública e epidemiologia não é recente e remonta aos tempos de uma Inglaterra vitoriana (SUSSER; SUSSER, 1996). Como conceitua a autora, para além do volume e da velocidade, o *Big data* tem propriedades importantes como variedade e valor. Essas propriedades se incorporam, hoje, à saúde pública e à epidemiologia moderna, ao considerarem dados para além das características clínicas e contextuais do paciente, tangenciando as manifestações de saúde e doença em contextos mais amplos como nas redes sociais.

Leonelli reconhece que "nem todos os bancos de dados são baseados em uma ideia clara de seus objetivos e estratégias de gerenciamento de dados" (p. 43), fator importante ao se considerar os Sistemas de Informação em Saúde (SIS) brasileiros. Criados em diversos momentos históricos para diferentes fins (em geral, não acadêmicos), sua utilização no contexto de *Big data* não pode ser ignorada desde que conhecidas as suas limitações e vieses. "[...] dúvida e desconhecimento sobre as razões e os critérios usados ao longo dos anos para escolher, adaptar e atualizar os métodos de gestão de dados" (p. 48) estabelecem o desafio para o trabalho com dados de SIS legados e pouco documentados.

A autora questiona situações e efeitos que podem ser observados na história dos SIS nacionais. Ao observar que as infraestruturas de *Big data* mais conhecidas estão "localizadas em lugares de poder no mundo científico" e são "estritamente mantidas e discutidas em inglês" (p. 60), é possível lembrar dos efeitos dessa relação até mesmo com dados textuais armazenados nos SIS: a inaptidão de boa parte desses sistemas para armazenar os chamados 'caracteres especiais' tão caros à língua portuguesa, como acentos e cedilhas.

Leonelli também lança luz sobre um tema de grande importância – e, por vezes, ignorado – na incorporação do *Big data*, que é a necessidade e o reconhecimento do poder da infraestrutura necessária para

a manutenção do *Big data*, que requer recorrentes investimentos financeiros para sua manutenção diante da expansão contínua dos dados e suas diversas propriedades. A convidativa alternativa de terceirizar essa infraestrutura, com a contratação de grandes empresas internacionais do ramo, requer reflexões sobre os conflitos de interesse gerados por suas posturas altamente comerciais de geração de valor (p. 67). É possível destacar que se observa, já na Introdução do livro, a importância dada pela autora ao valor comercial dos dados, independentemente de sua origem, se científica ou não (p.19), e à intrínseca relação do *Big data* com o mercado (p. 20). Inclusive, Leonelli destaca o receio revelado por grupos de pesquisa africanos de partilharem seus dados nessas infraestruturas internacionais e perderem controle sobre questões de autoria e ineditismo de pesquisa (p. 61).

Outro aspecto importante discutido pela autora é a "privatização dos dados" (p. 69). Na recente pandemia de covid-19, observaram-se movimentos em que os dados coletados e organizados pelo poder público foram fornecidos gratuitamente a empresas privadas que, por sua vez, revendiam para o próprio poder público o mesmo dado (VILLELA, 2021).

A proposta epistemológica relacional oferecida pela autora para se estudar *Big data* faz importantes contrastes com a abordagem representativa mais tradicional. Na abordagem relacional, Leonelli destaca que "os dados não são dados se existem apenas na mente de quem os usa: pelo menos em teoria, eles devem ser acessíveis a outros que possam avaliar seu valor científico e verificar sua confiabilidade como uma base empírica de conhecimento." (p. 98). É estranho não lembrar de situações em que dados potencialmente fabulosos, linkages extraordinárias de grande potencial para a saúde pública são propagandeados, mas nunca divulgados abertamente por seus autores.

Por fim, a autora faz sugestões para a incorporação de princípios à pesquisa com *Big data*, muito relevantes à pesquisa de saúde pública. As considerações sobre "integração da ética na pesquisa científica", "participação social" e a "necessidade de desacelerar os tempos de pesquisa" são temas caros e de contínua discussão na informação e comunicação em saúde e expandidos no livro.

O livro A pesquisa científica na era do Big data: cinco maneiras que mostram como o Big data prejudica a ciência, e como podemos salvá-la aadiciona importantes contribuições à biblioteca do pesquisador de informação e comunicação em saúde. A autora revisita e atualiza definições, convidando o leitor a fazer importantes reflexões e questionamentos sobre a utilização lúcida dos métodos de *Big data* na pesquisa científica.

REFERÊNCIAS

SALES, Luana Farias *et al*. (org.). **Princípios FAIR aplicados à gestão de dados de pesquisa.** Rio de Janeiro: Ibict, 2021. 292 p. (Coleção PPGCI 50 anos). DOI: https://www.doi.org/10.22477/9786589167242. Disponível em: https://ridi.ibict.br/handle/123456789/1182. Acesso em: 26 set. 2022.

SUSSER, Mervyn; SUSSER, Ezra. Choosing a future for epidemiology: I. Eras and paradigms. **American Journal of Public Health**, Washington, DC, v. 86, n. 5, p. 668-673, 1996. DOI: https://doi.org/10.2105/ajph.86.5.668. Disponível em: https://ajph.aphapublications.org/doi/abs/10.2105/AJPH.86.5.668. Acesso em: 26 set. 2022.

VILLELA, Daniel Antunes Maciel; GOMES, Marcelo Ferreira da Costa. O impacto da disponibilidade de dados e informação oportuna para a vigilância epidemiológica. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 38, n. 7, p. e00115122, 2022. DOI: https://doi.org/10.1590/0102-311XPT115122. Disponível em: https://www.scielo.br/j/csp/a/dDSpPy898L4Pj3WPLGxLdqN/?lang=pt. Acesso em: 26 set. 2022.