DOI: 10.1590/1413-81232020256.1.11792020

Personal data usage and privacy considerations in the COVID-19 global pandemic

Bethania de Araujo Almeida (https://orcid.org/0000-0001-8918-2661) [1]
Danilo Doneda (https://orcid.org/0000-0001-9535-3586) [2]
Maria Yury Ichihara (https://orcid.org/0000-0001-8590-6212) [1]
Manoel Barral-Netto (https://orcid.org/0000-0002-5823-7903) [1]
Gustavo Correa Matta (https://orcid.org/0000-0002-5422-2798) [3]
Elaine Teixeira Rabello (https://orcid.org/0000-0002-8324-1453) [4]
Fabio Castro Gouveia (https://orcid.org/0000-0002-0082-2392) [5]
Mauricio Barreto (https://orcid.org/0000-0002-0215-4930) [1]

Abstract *Data has become increasingly important and valuable for both scientists and health authorities searching for answers to the COVID-19 crisis. Due to difficulties in diagnosing this infection in populations around the world, initiatives supported by digital technologies are being developed by governments and private companies to enable the tracking of the public's symptoms, contacts and movements. Considering the current scenario, initiatives designed to support infection surveillance and monitoring are essential and necessary. Nonetheless, ethical, legal and technical questions abound regarding the amount and types of personal data being collected, processed, shared and used in the name of public health, as well as the concomitant or posterior use of this data. These challenges demonstrate the need for new models of responsible and transparent data and technology governance in efforts to control SARS-COV2, as well as in future public health emergencies.*

Key words *Personal data, COVID-19, Data governance, Technology governance, Public health emergency*

[1] Centro de Integração de Dados e Conhecimentos para Saúde, Fiocruz Bahia. R. Mundo, Trobogy. 41745-715 Salvador BA Brasil. baraujo2010@gmail.com
[2] Instituto Brasiliense de Direito Público. Brasília DF Brasil.
[3] Escola Nacional de Saúde Pública Sérgio Arouca, Fiocruz. Rio de Janeiro RJ Brasil.
[4] Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro. Rio de Janeiro RJ Brasil.
[5] Casa Oswaldo Cruz, Fiocruz. Rio de Janeiro RJ Brasil.

The growing production and use of data, made possible by increasingly powerful and specialized digital technologies, has empowered the emergence of new forms of knowledge production through sophisticated computational modeling and algorithms. In this new context, data becomes ever more important and valuable in a variety of contexts, including social, political and economic interests[1].

During the COVID-19 pandemic, the introduction of a previously unidentified etiological agent and the peculiarities of its accompanying disease present challenges and pose risks to the lives and health of populations worldwide, necessitating an urgent response. As a result, personal data from diverse sources has been requisitioned, under the presumption of ethical and legal usage, to investigate scientific questions based on populational characteristics, as well as data from laboratories and hospitals, among others.

A worldwide effort by scientists, organizations and health practitioners is being undertaken to close gaps in knowledge as quickly as possible to enable health authorities to introduce efficient clinical management and prevention measures to address the pandemic, including the agile implementation of improved diagnostic capacity and the rehabilitation of COVID-19 cases in a timely manner. These actions require articulation between governmental measures and different segments of society in order to maximize disease control efforts.

The WHO has advised that each country, in accordance with respective risk assessments, be prepared to respond to possible scenarios and rapidly implement necessary measures to reduce viral transmission and minimize economic and social impacts[2]. As a result, high quality data is needed to assess basic epidemiological patterns. Unfortunately, uncertainties surrounding COVID-19 also extend to the quality of data available to researchers, not only with respect to understanding underlying epidemiological patterns of disease, but also in the construction of mathematical models aimed at providing evidence to support decision making at diverse levels.

Considering the enormous burden posed by diagnosing infection in the general population, technological initiatives have been developed to enable the tracking of citizens' symptoms, contacts and movements, elements considered essential to the design of infection surveillance strategies by governments. Great hope lies in the development of applications that collect data on individuals, including their geolocation information and movements[3]. These practices raise questions regarding the type and amount of data required, and ethical, legal and technical challenges permeate related data collection, access, sharing and usage issues[4,5].

Apple and Google recently announced the joint development of a tool to track COVID-19 infection in a partnership aimed at ensuring interoperability between iOS and Android operating systems. According to the companies, users can opt in at their discretion, but there has been no mention of an option to subsequently withdraw consent. The tool, according to published specifications[6], bears similarities to other contact tracing solutions, broadly inspired by those already in operation in Singapore and proposals under development in Europe, such as DP-3T (*Decentralized Privacy-Preserving Proximity Tracing*)[7] or the PEPP-PT (*Pan-European Privacy-Preserving Proximity Tracing*)[8] project and MIT Safe Paths Platform, which seek to maximize privacy[9].

These mobile system solutions, which can be broadly classified as contact tracing systems, generally function through the short-range exchange of anonymous identifiers via Bluetooth technology. Depending on the solution, an application made available by national health authorities can be installed, or the technology may eventually be "baked-into" operating systems. Users who receive positive coronavirus test results register their status in the application, which then communicates this to respective health authorities; others with whom the user came into close contact during the previous 14 days will also receive alerts[10]. As these are technologies still under development and undergoing maturation, differences in implementation could, over time, prove very significant; as an example, consider what appears to be the centralized focus of the PEPP-PT compared to the decentralized approach of DP-3T.

The current panorama surrounding the Coronavirus epidemic indicates that, during the next phases in which society will continue adapting to living with the virus, the use of personal data and applications or devices will play a prominent role not only in gauging contact, but also for purposes such as verifying the compliance of isolation or quarantine measures, which may extend to probabilistic contagion verification or managing permissions for citizens to go out in public, among many other uses.

It is important to remember that data collection through applications and smartphones requires access to these technologies and users

Ciência & Saúde Coletiva, 25(Supl.1):2487-2492, 2020

must necessarily be familiar with their usage; this implies that the data collected will be representative of certain populational groups. Accordingly, adopted measures must consider health inequalities and accommodate differences in the impacts of solutions on diverse segments of populations.

In addition to location tracking, encouraging users to self-report symptoms and automatically sending alerts about possible contact with infected individuals, personal data, such as patient health information, is being used in other ways. In the UK, for example, government agencies have been working with technology companies to build a COVID-19 repository containing patient data. These companies were hired by the National Health Service (NHS) to assist with the elaboration of predictive models using artificial intelligence and patient data. The initiative was justified by the need for information regarding the burden on health services in real time using hospitalization data and intensive care bed availability, as well as equipment and supply needs.

The NHS has declared that the data in this repository is confidential, anonymized and stored in a government database, and that it will remain under its control and subject to severe restrictions under data protection legislation; nonetheless, the initiative has aroused the public's mistrust regarding ethical, privacy and data protection aspects of these citizens' private information[11].

Questions and challenges have been raised regarding the public's trust in the institutions, whether governmental or private, responsible for processing personal data. This wariness and questioning does not aim to prevent the use of data in the response to the pandemic, but highlights the need to establish safeguards to ensure a balance between individual and collective interests as well as to increase societal confidence in the institutions processing data for public health purposes[12,13].

In Brazil, the General Data Protection Law (LGPD), which was approved and sanctioned in 2018, is scheduled to take effect in August 2020; however, this could change as bills currently under consideration by the Brazilian congress seek to delay adoption until 2021. LGPD represents a milestone in the regulation of personal data, since it applies to all personal data handling operations, including in the arena of digital media, whether by individuals or public and private companies. This law was devised to protect the fundamental rights of freedom and privacy[14].

Informed self-determination is undoubtedly a fundamental aspect to be taken into consideration regarding the use of personal data, together with guarantees of transparency, security and the minimization of data usage. However, in the case of emergency situations and others in the public interest, such as a public health crisis, the use of personal data is allowed in the absence of citizens' consent, provided that safeguards are put in place, the data is used precisely to achieve specified purposes and the agencies authorized to process data are qualified in accordance with regulations established in the Brazilian General Data Protection Law and the General Data Protection Regulation adopted by the European Union[15]. Several elements of the LGPD are designed to enable the use of personal data in policymaking and systems formulated to combat COVID-19, which can be used as soon as the law goes into force[16].

Anonymization, which consists of applying technical measures to render the direct or indirect association of data with a given individual impossible, and pseudo-anonymization, which generally removes and replaces identifiers with a unique key code, are examples of data protection strategies incorporated into law. With few exceptions, anonymized data is generally not considered personal data, while pseudo-anonymized data is still considered personal data due to the potential for reidentification of individuals through a key code, potentially high levels of security. Due to the possibility of identifying anonymized data[17], combinations of various procedures are necessary to preserve individual privacy, particularly when databases are integrated[18].

Compliance with general data protection laws, therefore, requires technology, infrastructure and specialized personnel to ensure that personal data are processed in a lawful, fair and responsible manner. Moreover, accountability must be guaranteed through the monitoring of data processing activities by designated authorities authorized to apply sanctions in the case of transgressions. In some countries, partnerships between government, universities and research institutes have created data centers to process and provide access to anonymized data in a secure and controlled manner to support investigative research in the public interest[19].

Anonymized or aggregated data are not considered personal data by data protection laws, since the identification of individuals is protected. However, even without referring to any specific individuals, groups could still be harmed due to the aggregation of information on locale, ethnicity, health situations and socioeconomic conditions, necessitating ethical scrutiny regard-

ing the potential benefits generated by such evidence.

Linnet Taylor has called attention to the fact that no protection exists against irresponsible technologies, as data protection laws focus exclusively on the protection of personal data, yet do not cover the freedoms and political rights of collective groups. Civil society groups must be allowed to participate in the governance of technologies. She argues that technology companies must be transparent and accountable to society in order to validate their legitimacy as actors on behalf of the government and population, at least those companies that, in light of the pandemic, have partnered with governments and now participate in the governance of citizens' data[20].

Considering that data can be used and shared by different people and organizations simultaneously, the main issues that need to be addressed pertain to responsible data governance based on transparency and citizen empowerment to fortify trust and establish balanced and fair relationships between individuals and organizations[21].

The legitimacy of collecting, processing, sharing and using personal data does not come from access to this data, but rather from trust in whomever possesses it, treats it with transparency and operates within legal parameters. From this perspective, the use of personal data to face COVID-19 and future public health emergencies must be guided by transparency, verification and accountability, beginning with collection and extending onwards to processing operations and the purported use of data, as well as by whom and for how long[22].

Clear and transparent terms and conditions must be applied regarding the access, sharing and use of the data collected in the name of public health, especially by private companies or through public-private partnerships. How and by whom will this data be accessed, processed and used? Will the data be stored, reused or discarded after the initial objectives are achieved? How will the data be protected? In the case of abuse or neglect, who will be held responsible? These and other questions should be asked and answered explicitly.

Another regulatory aspect that deserves attention pertains to intellectual property rights, as the selection, organization or availability of data stored inside databases is protected by intellectual property rights[23]. Databases that can be integrated with data from other sources to subsidize the development of new technologies, including treatment and prevention technologies for COVID-19, will be subject to ownership rights and could potentially incur costs related to access.

Partnerships between governments, technology companies and universities are necessary to enable the extraction of reliable knowledge from large volumes of data. The agreements covering these ventures must clearly specify the roles of the parties involved, as well as usage of the presumed and achieved results. The establishment of protocols with guiding principles providing for the agile and practical application of data processing in cases of collective interest, such as the current health emergency, is urgently needed, especially considering the national and transnational utilization of personal data collected by companies around the world.

Responsible data governance also entails the description of data processing and analysis methodologies, as data can be provided as proof, as evidence, in decision making for both public policy and science[24]. Importantly, any machine learning-based algorithm is representative of the pattern or regularity of what it was intended to measure. Algorithms are powerful and important resources that cannot be separated from causal explanations due to the risks of making decisions based only on automated results and predictions. The predominant role of the scientific method is thusly to validate and increase the reliability and usefulness of results. Indeed, science is preoccupied with the questioning of assumptions, values and biases in order to distinguish opinions from evidence.

Regulations are the only mechanism capable of establishing limits on the processing of personal data by governments and private corporations, even in a health crisis, to avoid negative impacts resulting from temporary relaxation, which have the potential to become permanent, as was seen in the United States following the September 11th attacks in 2001. Public surveillance strategies were introduced using existing and emerging technologies at the time, justified by the monitoring of suspicious individuals and in order to avoid future terrorist aggression, resulting in lawful changes arising from fear instilled in society[25].

The adoption of more just, responsible and sustainable data governance models, designed to protect and defend ethical and regulatory principles, serves to increase the confidence of individuals and society in the use of personal data to respond to situations of legitimate public interest. Aspects related to the privacy rights, the protection of personal data and the rights of groups do

Ciência & Saúde Coletiva, 25(Supl.1):2487-2492, 2020

not preclude the use of personal data, especially in response to a pandemic. The public health emergency provoked by SARS-COV-2 highlights the pressing need for new forms of personal data governance that include civil society, with the goal of promoting equitable benefits for society as a whole.

Collaborations

BA Almeida conceived and led the writing of the first version of this article, and also contributed to the editing and revision of the final text. M.L Barreto supported the design and writing of the original article and also contributed to the editing and revision of the submitted text. The other authors, D Doneda, MY Ichihara, M Barral-Netto, G Matta, E Rabello and F Gouveia supported the writing of the original article and were also responsible for editing and revising the text. The authors would like to thank AK Walter for English language revision and copyediting services.

References

1. Leonelli S. Data – from objects to assets. *Nature* 2019; 574:317-320.
2. World Health Organization (WHO). *Critical preparedness, readiness and response actions for COVID-19*. Geneva: WHO; 2020.
3. The Economist. *Covid-19. App-based contact tracing may help countries get out of lockdown but only as part of a bigger system*. [acessado 2020 Abr 16]. Disponível em: https://www.economist.com/science-and-technology/2020/04/16/app-based-contact-tracing-may-help-countries-get-out-of-lockdown
4. Kim MJ, Denyer S. A 'travel log' of the times in South Korea: Mapping the movements of coronavirus carriers. *The Washington Post*. [acessado 2020 Abr 16]. https://www.washingtonpost.com/world/asia_pacific/coronavirus-south-korea-tracking-apps/2020/03/13/2bed568e-5fac-11ea-ac50-18701e14e06d_story.html
5. Gilbert D. Iran Launched an App That Claimed to Diagnose Coronavirus. Instead, It Collected Location Data on Millions of People. *Vice News*. [acessado 2020 Mar 14]. Disponível em: https://www.vice.com/en_us/article/epgkmz/iran-launched-an-app-that-claimed-to-diagnose-coronavirus-instead-it-collected-location-data-on-millions-of-people
6. Newsroom. *Apple e Google formam parceria para tecnologia de rastreamento de contato com COVID-19*. [acessado 2020 Abr 10]. Disponível em: https://www.apple.com/br/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/
7. Payer M, Barman L. *Descentralized Privacy-Preserving Proximity Tracing*. [acessado 2020 Abr 10]. Disponível em: https://github.com/DP-3T
8. *Pan-European Privacy-Preserving Proximity Tracing*. Disponível em: [acessado 2020 Abr 10]. https://www.pepp-pt.org/
9. Project Safe Paths. *Massachusetts Institute of Technology*. [acessado 2020 Abr 10]. Disponível em: https://www.media.mit.edu/projects/safepaths/overview/
10. Panzzarino, M. Apple and Google are launching a joint COVID-19 tracing tool for iOS and Android. *Tech Crunch*. [acessado 2020 Abr 10]. Disponível em: https://techcrunch.com/2020/04/10/apple-and-google-are-launching-a-joint-covid-19-tracing-tool/
11. Lewis P, Conn D, Pegg D. UK government using confidential patient data in coronavirus response. *The Guardian*. [acessado 2020 Abr 12]. Disponível em: https://amp.theguardian.com/world/2020/apr/12/uk-government-using-confidential-patient-data-in-coronavirus-response?CMP=share_btn_tw&__twitter_impression=true
12. European Digital Rights. *EDRI calls for fundamental rights – based responses to COVID-19*. [acessado 2020 Mar 20]. Disponível em: https://edri.org/covid19-edri-coronavirus-fundamentalrights/
13. Mcknight G. Coronavirus surveillance concerns ramp up pressure to pass federal privacy law. *Internet Governance Hub*. [acessado 2020 Abr 10]. Disponível em: https://www.internetgovernancehub.blog/2020/04/10/coronavirus-surveillance-concerns-ramp-up-pressure-to-pass-federal-privacy-law/
14. Brasil. Lei n. 13.079, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). *Diário Oficial da União* 2018; 15 ago.
15. European Union (EU). *Regulamento Geral de Proteção de Dados da União Europeia – EU 2016/679 (GDPR)*. [acessado 2020 Mar 20]. Disponível em: https://gdpr-info.eu/
16. Doneda D. *Opinião e Análise. A proteção de dados em tempos de coronavírus*. [acessado 2020 Mar 20]. Disponível em: https://www.jota.info/opiniao-e-analise/artigos/a-protecao-de-dados-em-tempos-de-coronavirus-25032020
17. Rocher L, Hendrickx JM, De Montjoye YA. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun* 2019; 10:3069.
18. Harron K, Dibben D, Boyd J, Hjern A, Azimaee M, Barreto M, Goldstein H. Challenges in administrative data linkage for research. *Big Data&Society* 2017; 4(2):11-12.
19. Doneda D, Almeida BA, Barreto ML. Uso e proteção de dados pessoais na pesquisa científica. *Revista Direito Público* 2019; 16(90):179-194.
20. Nuffield Council on Bioethics and Ada Lovelace Institute. *Webinar - Beyound the exit strategy: ethical uses of data-driven technology in the fight against COVID-19*. [acessado 2020 Abr 20]. Disponível em: https://www.nuffieldbioethics.org/publications/covid-19/webinar-beyond-the-exit-strategy-ethical-uses-of-data-driven-technology-in-the-fight-against-covid-19
21. My Data Global Blog. *My Data vs COVID-19*. [acessado 2020 Abr 20]. Disponível em: https://mydata.org/2020/04/06/an-approach-for-fighting-covid-19-and-beyond-mydata/
22. Patel R. Removing the pump handle - stewarding data at times of public health emergency. [acessado 2020 Abr 8]. Disponível em: https://www.adalovelaceinstitute.org/removing-the-pump-handle-stewarding-data-at-times-of-public-health-emergency/
23. Guanaes P, Souza AR, Doneda D, Nascimento FJT. *Marcos legais nacionais em face da abertura de dados para pesquisa em saúde: Dados pessoais, sensíveis ou sigilosos e propriedade intelectual*. Rio de Janeiro: Fiocruz; 2018.
24. Leonelli S. Data Governance is Key to Interpretation: Reconceptualizing Data in Data Science. *Harvard Data Science Review* 2019. [acessado 2020 Abr 8]. Disponível em: https://hdsr.mitpress.mit.edu/pub/4ovhpe3v
25. Levi M, Wall DS. Technologies, Security, and Privacy in the Post-9/11 European Information Soci*ety. Journal of Law and Society* 2004; 31(2):194-220.

Article submitted 27/04/2020
Approved 29/04/2020
Final version submitted 01/05/2020