

The challenges of urban data ethics: big data and intelligent decision support systems

Larissa Galdino de Magalhães Santos¹

¹ Getulio Vargas Foundation, Law School, Rio de Janeiro, Brazil
larissagms@yahoo.com.br

Abstract. Transparency is the main format of any initiative that uses data. When it comes to sensitive data and big data to support decision making, the combination can bring advances but also additional concerns in the field of data ethics. This article examines the Intelligent Monitoring and Information System (SIMI-SP), investigating the creation of the isolation index, made available and maintained by telecommunication service providers (telecoms), through a big data platform. It offers a picture of how this data-oriented technology is being instrumented, combined with other data, generating evidence that can contribute to the management of the pandemic and increase the risks of privacy, data protection and security. Finally, it highlights the potential of technologies in relation to urban data ethics, the challenges faced and possible outputs. Keywords: open data, big data, decision making, ethics.

Keywords: open data, big data, decision making, ethics.

1. Introduction

The growing use of urban data impacts policy and government planning towards smart governments and cities. With the smart city, data-driven urban governance now influences public policy formulation as well as service delivery. The lines of the physical world and cyberspace are blurring the distinction in public policy areas. Consequently, urban data are based on a computational understanding of cities, whose logical procedures, automated systems and algorithms, all based on instrumental rationality, now inform and sustain the generation of urban information.

What happens when the smart city encounters disasters and pandemics on a large scale like COVID-19 and needs to improve decision making? There is an effort to respond to the pandemic [1], whether by science, governments, private sectors, or

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

social organizations, and that passes through the capacity of data governance capable of articulating measures to control the spread of the virus, even the impacts post-pandemic.

Therefore, the pandemic has raised important questions about how to open data, use and share data, in the face of the continuous need to make decisions. Faced with the need for quick responses, the State Government of São Paulo promoted the creation of an intelligent system, including the creation of an isolation index based on data from telecommunications companies. To meet these demands, challenges are placed throughout the data value chain, in a scenario that involves interests and negotiations of multiple dimensions and that gains more relevance, the governance of data.

If on the one hand the governance of data agglutinates ways to explore, organize and regulate the information and data that are invading the social fabric of urban life. On the other hand, data governance manages from usability, integrity, security and data use patterns, which imply concerns about privacy, security, protection and ethics. Ethics, in the context of intelligent cities, and open urban data are not strongly debated, nor are they combined with the invasive concerns of technology [2].

In fact, the lack of appropriate regulations and control mechanisms can drive technologies and their results to society in an imposed and incorrect way. The pandemic has reinforced the critical context by laws, mitigation measures, and ethical codes concerning data analysis and its operationalization.

Thus, there are gaps in how digital technologies combined with data and converted into intelligent and automated systems, algorithms and applications, generate evidence and decisions that may prevail over the scrutiny of public institutions, knowledge and transparency of the citizen. And, yet, on the reconfiguration of public habits, civic trust and transformations through a data-driven urbanism order.

This article has an objective, to examine the Intelligent Monitoring and Information System (SIMI-SP), investigating the creation of the isolation index, made available and maintained by telecommunication service providers (telecoms), through a big data platform. Plus, it highlights the potential of the technologies in relation to the ethics of urban data, the challenges faced and possible outputs.

I argue that these crises, besides being devastating, are a turning point in the urgency of urban data ethics. In general terms, data processing technologies involve choices, and choices involve values, norms and codes. At the same time, the speed, scale, and diffusion of data reveal advances with impacts that are difficult to predict, regulate, reformulate, and execute. Communities are being driven to reflect on the challenges of data ethics.

2. The challenges of urban data ethics

Recent research in the fields of computer science, computer social science, statistics, geospatial technologies, and machine learning domains reflect the optimism associated with the use of tools and voluminous data to find and process accurate in-

formation capable of driving rapid decision making and providing effective solutions to social emergencies.

Data is processed to obtain information and perform tasks, and that influence aspects of the city - social, economic, cultural, reveal concerns about conduct and ethics. In these projects, the ethical and safe use of urban data should be a priority [2].

In smart cities, digital technologies have evolved into an extensive ecosystem of software, services, sensors and networks, whose data from a variety of sources can lead to sensitive or private information. Thus, urban data top the list of promises for the most efficient and effective public policy responses. If technologies promote social advances and quality of life improvements, they also carry concerns, vulnerability and challenges of privacy, security and protection [3].

This analysis is based on the assumption that decision making based on the lack of "raw, real and qualified" data, biased neural networks or weak governance models, may neglect the lines of ethical conduct drawn in the urban landscape, triggering higher survival costs of surveillance, control and protection.

Technologies are not ethically 'neutral' and need to establish ethical conditions and prerequisites. Technologies involve choices and choices involve values, standards and ethic-driven codes. The rules of the "game" have changed and there is a strong imbalance in the balances of power. The meeting of these trends in the Covid-19 scenario in the State of São Paulo has intensified the questions about how governments and multiple actors can overcome their problems and achieve solutions, managed and integrating data, information, technologies and services in an intelligent way, either to generate evidence in public policies or to promote collaborative solutions among all actors in the territory.

Thus, the ethical challenges of urban data are on four main levels: i. platform capitalism; ii. the order of socio-digital inequality; iii. intelligent governance in the context of urban data; iv. flood of urban data and the scientific approach.

In a broad sense, for example, platform capitalism [4] or data capitalism [5] projects the business models that use, extract and monitor user and device data as a source of power in cities, and has a strong impact on privacy boundaries, geographic surveillance and social classification, the private sphere, urban life, and thus data ethics.

The current socio-digital [4] order can generate deficient digital capital, as there is an unequal distribution of resources that facilitate access to Internet access infrastructure and the digital skills to benefit from the network. The digital capital deficit is driven and reinforced by structural inequalities. Groups marginalized and digitally excluded continue to survive within the margins imposed by socio-digital orders. Artificial intelligence (AI) systems "produce" political evidence projected from fingerprints or even misuse, increasing the risk of reproduction and perpetuation inequalities.

Smart governance requires reshaping the role of government, citizens and multiple actors combined with technology and resource tools. Only technologies and data analysis are not able to sustain the links and complexity of the city's social interac-

tions. The citizen needs a "voice" and "vision" of the policy formulation and decision-making process. There is a discrepancy between data governance and smart governance projects [4] that fail to activate and interoperate data and cyber-physical devices in spaces marked by unequal access and infrastructure of Internet and digital technologies.

The processing and analysis of data that is of high relevance and responsibility, such as learning machines to handle huge data sets, has an epistemological approach that is realistic and objective [6], that can ignore aspects of human life related to social structures, cultures and ideology, and still produce failed interventions, causing damage to cities.

The decision making and public policy, based on big data inferences or urban data, can neglect the complex lines drawn in the urban landscape. Consequently, decisions and policies can trigger higher survival costs in areas already marked by social inequality.

3. The São Paulo Government's Isolation Index and Intelligent Monitoring and Information System (SIMI-SP)

The State of São Paulo has developed partnerships with operators, the Secretariat of Economic Development and the Institute of Technological Research, to use data related to mobile telephony and the displacement of users, and ways to contain the virus and formulate policies.

In March, State Decree n° 64.86429ⁱ defined the prerogatives of the Extraordinary Administrative Committee COVID-19. The partnership with the operator Vivo of the company Telefônica Brasil S. A., was celebrated under the terms of the technical cooperationⁱⁱ that regulated the form of access to information provided by the Big Data Platform of the telecommunication company. The agreement regulated the rules of secrecy and confidentiality of information, making available visualization maps with anonymized and aggregated data.

In the second stage, the Intelligent Information and Monitoring System (SIMI-SP) was updated with a new technical cooperation with the main telephone operatorsⁱⁱⁱ, Claro S.A; Oi Móvel S.A., Telefônica Brasil S.A., Tim S.A. and through the Brazilian Association of Telecommunication Resources - ABR - Telecom. The agreement allowed the institute access to anonymized data through heat maps available on the Big Data Platform of telecoms.

The isolation index is one of SIMI's resources, and the maintenance of the panel, in its computational environment is the responsibility of the Institute of Technological

ⁱ<https://www.al.sp.gov.br/repositorio/legislacao/decreto/2020/decreto-64864-16.03.2020.html>

ⁱⁱ<https://www.ipt.br/download.php?filename=1928->

[Extrato_Termo_de_Cooperacao_Tecnica_Telefonica.pdf](#). Accessed 2020/06/10.

ⁱⁱⁱ<https://www.ipt.br/download.php?filename=1920->

[to_ACT_Prestadoras_de_Servicos_de_Telecomunicacoes.pdf](#). Accessed 2020/06/10.

Research with that of the Management Committee. The social isolation index, fruit of the second technical cooperation agreement with telecoms and ABR Telecom, the data are obtained as follows: first, the operators process the mobility data in the cell phone terminals, the Base Stations; after processing, the operators release the access to the platform interface, for the IPT to visualize the index, graphs and maps; the institute makes the data available for publication in the SIMI panel.

As the index is updated daily, it has as reference the values of the previous day. According to the IPT^{iv}, in this period the operators aggregate and anonymize the data to generate the indexes that will be made available to SIMI. The index is based on the location of cell phone antennas, that is, Base Radio Stations, so that it is possible to indicate periods of permanence of cell phones the art of a reference location. According to the Brazilian Telecommunications Association, in the state of São Paulo, there are about 25.230 antennas.

They are related to these stations that are processed by the operators. Obtaining the information at the Base Station is independent of the triggering of a GPS device or other location application, because the data comes from the cell phone identification when it is used for calls or to access the data. These services depend on the provision of data through radio frequency.

The IPT (Technological Research Institute) reinforced that no personal data is passed on to the system. Emphasizing that "the operators are responsible for the information security of their databases, applying the premises of the General Law of Data Protection in the context of the cooperation agreement signed with IPT^v". The institute also emphasized that the data of the telecoms platform has no relation with the information on health records.

5. Discussion

The main companies in the telecommunications sector in Brazil have created a data platform for locating and moving cell phone users. Through the communication and radiofrequency stations of the telephony provision and data exchange services, the support entities, SindiTelebrasil (National Union of Telecommunications Companies) and ABR Telecom (Brazilian Telecommunications Resources Association), have joined efforts to create a data lake with statistical and unidentified data, to foster political decisions and the management of the pandemic.

The State Government of São Paulo was a pioneer in creating intelligent systems to support decision-making in pandemic management. The case study on the Information System and Intelligent Monitoring highlighted the stages of updating technology and interactions with telecoms through a cooperation agreement. The big data platform has also been improved, in volume and data accuracy, extending the scope of the heat maps for both the insulation index and the new agglomeration index.

^{iv}.<http://www.telecocare.com.br/telebrasil/erbs/> Accessed 2020/06/11.

^v.<https://www.saopaulo.sp.gov.br/coronavirus/isolamento/> Accessed 2020/06/11.

However, this is a political process that highlights a series of provocations about surveillance, control and accuracy of data. Certainly, the technology offered freely by telecoms has strong potential to support public policies, especially in the pandemic context. Questions about the violation of rights and violation of privacy in the case of SIMI-SP, were taken to the organs of the State Court of Justice, in addition to warrants of safety of users who felt harmed.

Demonstrations on the big data platform in response to COVID-19, addressed criticism of the risk of the tools facilitating discrimination or abusive use, the need for a public commitment to the use of data^{vi}.

These challenges are associated with the four levels of research analysis: platform capitalism and the instrumentalization of digital tracks that are fed by algorithms indicating patterns and behaviors; socio-digital inequality that can undermine the planning and management of policies of those on the margins of the digital universe and that requires mitigation and inclusion measures; intelligent governance in the urban context linked to the flood of data and intelligent systems reinforce the responsiveness and transparency of public actions and decision making.

Platform capitalism, data instrumentation and algorithms As the social world becomes increasingly active online, it is likely that monopoly, data control combined with inequalities and social inequalities will undermine attempts to use data to improve citizens' quality of life. Large platforms now use application data, full of rich predictive digital traits, to arbitrarily monetize personal data. This information is used to exploit individualities, focusing on behavior, and compromising rights in a liberal economic scenario [7].

The big data platform with heat maps of the telecoms is not accessible to public knowledge, neither are the documents of the cooperation. It is not possible to evaluate the data collection and storage policy. Although the entities have expressed their commitment to the General Data Protection Act, the declarations do not guarantee consistency regarding the process of data collection, analysis and availability. The collection of all of them seems algorithmically confusing.

Although the register in the Radio Base Station is a statistical data, according to SindiTelebrasil, there is an identification of the lines that are connected in a certain station, by day, hour and minute, and there is no knowledge about the type of processing on this large volume of data. I argue that lack of clarity with the method impacts on the use of data, reuse, and the use of information. The risk reinforces the lack of knowledge about the techniques and modeling of statistical data, clusters and anonymized by operators. And also how inferences about big data can be used in other situations.

Fostering explanations and transparency about how people's data will be used for other purposes can ensure public trust. The posture relegates the traces of data capital-

^{vi}vi. <https://www.convergenciadigital.com.br/cgi/cgilua.exe/sys/start.htm?UserActiveTemplate=site&UserActiveTemplate=mobile&infoid=53275&sid=8>. Accessed 2020/06/15.

ism, in that the distribution of asymmetrical power is the result of the access and capacity of understanding to data processing by a select set of actors [8].

The narrative justified by the use of technologies in community benefits frames modes of surveillance that are barely perceptible because it produces its conceptions of authority and power through digital traces. It is a narrative of normalization of authority. This is not an observation of harmful consequences from the big data platform of telecoms, but a warning of how initiatives can take forms of algorithmic discrimination, influences in the domains of everyday life.

In the case of SIMI-SP, measures were taken to clarify the cooperation and partnership, however, the government did not take the initiative to present details on data processing and methods, delegating full responsibility to the telecoms and entities that assumed the cooperation. SIMI-SP, offers a complete bulletin of pandemic data, accompanied by a document with the methodology used to collect and process data related only to the public health panorama in the State.

The narratives based on the social and political benefits of the technologies and big data that surround these initiatives, contribute to justify the lack of transparency in view of the networks' capacity to promote impacts in the community. Opening the entire procedure from the collection to the delivery of data on the big data platform can build trust and commitment, and activate projects of change and transformation both private and public sectors.

The key is to increase impact, break down the boundaries of organizations and sectors, because when the procedure is better informed, it can also be better tested, gathering peer review, achieving greater convergence and multi-stakeholder efforts.

Therefore, sharing the journey should be the pattern, even more so with regard to the pandemic, when everyone is dealing with the same challenges. Opening codes, data and prototypes can improve data collection and storage policies. In addition, good data begins with a consistent mode of governance. Able to aggregate in the creation and operation of entities that promote the development and use of data in conditions of progress.

Experiences such as the platform and SIMI can highlight the importance of an authority, a movement, a code, that promotes the maintenance of the data lifecycle, connects different actors, centralizes parameters and encourages social participation and transparency in any project involving open data. To do this, creating a roadmap to define the data infrastructure is the first step. Followed by a program of standards, terminologies, workflows, documentation, to fill gaps to support evidence-gathering procedures and decision-making.

Poorly structured governance models can ignite a spiral of civic distrust. SIMI-SP has been imposed in court on the scope of data surveillance and the conditions of the agreement. There are reports that the data began to be used before a formal cooperation agreement^{vii}.

^{vii} <https://olhardigital.com.br/noticia/justica-libera-vigilancia-de-celulares-em-sp-para-controle-de-isolamento/101842>. Accessed 2020/06/15.

The system was considered safe and the use of the data in accordance with the law. However, the judicial manifestations confirm the lack of explainability of the platform's procedures and the forms of appropriation of information by governments. It is expected that the intelligent government will be able to integrate the use of technologies in favor of a collaborative environment with the citizen, in transparent and responsible actions.

However, if there is no understanding of the decisions and actions of the public sector based on data, it is unlikely that the citizen will be able to audit, supervise the government appropriately. Therefore, there is no way to ensure that decisions are accountable in management, monitoring and execution.

When the State Government of São Paulo emphasizes that the platform data are entirely the responsibility of the telecoms and entity, it is retreating to the opportunity of opening the government based on a structure of civic trust. Intelligent, citizen-centric government must put the genuine user experience first to create a dignified experience for citizens when they interact with government.

Citizen-centric services require good data and metadata, including geospatial information on government services and the eligibility and calculation rules related to these services. Constant feedback loops involving citizen input, ideas and experiences are extremely important to ensure iterative improvements over time to keep services relevant and responsive to changing population needs. No verification and feedback measures have been taken to citizens and civil society organizations if decisions made on the basis of the information from the big data platform have reached tangible levels of isolation. Obtaining the information does not guarantee that good decisions will be made. Between new and old partners and cooperation between telecoms, entities and governments, there is no kind of evaluation if the information is providing refined analysis. Still, it is not clear if there is any secondary use of data and information, either for governments or companies.

The ultimate implication of this wire is that data monopoly or procedures related to the data lifecycle, from collection to processing, are resources almost unique to data scientists, companies and platforms. These unique resources, growing volume of data, sensors and devices, are not "common vocabulary" for public or social agents.

Therefore, the appropriation of information generated by this large database, or analyses generated in the field of computer science, which lack "translation", transparency, objectivity and responsiveness. Open algorithms, open rules are necessary labels to guarantee the decisions and rights of users.

This means that local and state initiatives such as SIMI-SP should "force" the big data platform to design the explainability of methods and decisions for data collection and maintenance in the system, details about the use of artificial intelligence, and the possibility of data aggregation.

Otherwise, the intelligent system becomes a black box, which is inappropriate for the public sector and for any activities within the digital transformation. Any rule I may diminish transparency or compression should be brought to the debate. The man-

agement of the big data platform and the respective cooperations can encourage the applicability of rules that create gaps in accountability to users and citizens.

Without privacy and protection, data ethics is emptied. How important is data ethics in this scenario? Countries, such as Brazil, where the organization of a data authority is crawling, together with a context of misinformation and fake news, a data ethics movement has much to promote. It is hoped that this article has provided some clues for reflection on data ethics, and can shed light on programs, policies, and services that operate on the basis of civic trust.

Actually, choices in data ethics have been guided and prioritized by political and economic actors capable of changing, controlling and influencing other social, legal, institutional and technological aspects of digital infrastructure. In this analysis, the big data platform of telecoms and SIMI-SP was no different. Political and economic actors as protagonists of data governance formats.

When these political and economic actors are committed to poorly structured initiatives or governance models without transparency and responsiveness, ethical resources pale and the results are: lack of transparency, privacy, openness of government, no investment in digital skills, obscure data and controversial information combined with unexplainable processing methodologies.

This can reinforce vulnerabilities, systems and algorithms that lack methodological rigor and do not translate reality. It is a cycle of ethical problems and to modify and reformulate data ethics, it is necessary to activate technical resources to organize the entire data infrastructure; cultural resources to promote professional care and data appropriation; and regulatory resources to ensure regulations focused on privacy and data protection at individual and collective levels. How to proceed? How to guarantee the ethical conditions? How to guarantee the citizen's privacy? Spread and improve the technical knowledge of all those involved with the data.

Ensure storage security for data residues. Facilitate and encourage the participation of multiple stakeholders in the ecosystem, and in procedures, standards, concepts and categories for updating data and platforms.

Promote the use of non-proprietary software. Promote the creation of an Internet civil rights framework, approaches to prevent news falsification and misinformation, such as a data ethics protocol. Align with the transparency that is the primary format of any open data initiative and entity.

Promote accountability rules in data use and experience sharing. Reinforce existing initiatives, as the expansion of data, devices and technologies will not be interrupted. However, the damage risks and effects have continued to arise, and measures must be established to resist the impact of their problems.

References

1. Almeida, B. A. et al. Preservation of privacy when facing COVID-19: personal data and the global pandemic. *Ciênc. saúde coletiva*, Rio de Janeiro, v. 25, supl. 1, pp. 2487-2492, jun. (2020).
2. Clever, Sawyer et. al. Ethical Analyses of Smart City Applications. *Urban Sci.* 2, no. 4: 96. (2018).
3. Kitchin, R. The ethics of smart cities and urban science. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 374(2083), (2016).
4. Beaunoyer, E. et al. COVID-19 and digital inequalities: Reciprocal impacts and mitigation strategies. In: *Computers in Human Behavior* 111 (2020).
5. Srnicek, Nick. The challenges of platform capitalism: Understanding the logic of a new business model. *Juncture.* 23. ppp. 254-257. (2017).
6. Pham, Q et. al. Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts. Preprints (2020).
7. Couldry, Nick; Mejias, Ulises. Data colonialism: rethinking big data's relation to the contemporary subject. *Television and New Media.* ISSN 1527-4764 (In Press). (2018).
8. Arvidsson, Adam. Facebook and Finance: On the Social Logic of the Derivative. *Theory Culture & Society* 33, no. 6, pp. 3-23, (2016).