ARTHA
GLOBAL

Working Paper 01

The State of Open Data in Urban India

Harsh Vardhan Pachisia
Anushka Bhansali and
Rajeswari Parasa

December 2022

artha.global

About Data Governance Network

About Data Governance Network

The Data Governance Network is a self-sustaining, multidisciplinary community of experts which bring institutions, individual experts, and stakeholders from industry, government, and academia together through its data-backed research, policy discussions, and convening activities. This results in insightful and varied knowledge exchange contributing to the enhancement of technology-focused policy research

About Artha Global

Artha Global is a research, consulting and network facilitation organisation that assists governments in the developing world to design, implement and institutionalise policy frameworks that promote freedom, prosperity, and stability for citizens in the 21st century. We provide actionable research and support policy implementation and institutionalisation. We also create convening platforms to conceptualise new development agendas and build cross-sectoral coalitions across government, business, academia, philanthropy and civil society. Our work focuses on the structural transitions that developing countries must undergo to achieve sustainable and inclusive development. These economic, spatial, social and technological transitions can be summarised as the move from rural to urban, informal to formal, farm to factory, analog to digital, and low to high productivity. Additionally, they take place against the backdrop of new imperatives like climate change, gender equity and data privacy, and shocks like pandemics. We help governments build the requisite capacity to manage these transitions (and the inevitable dislocations they cause), in order to secure long-term prosperity and social stability for their citizens.

Acknowledgements

We would like to thank the Data Governance Network's review team, Bianca Shah, Jagan Shah, and Abhay Pethe for reviewing and providing feedback on the paper. We are grateful to Vikram Sinha, Hemant Adarkar, Avanti Durani and Sridhar Ganapathy for providing valuable insights as well as all participants at the Data Governance Network's Quarterly Roundtable on 20 May, 2022 where this paper was presented. This study has benefited from conversations with Niranjan Rajadhyaksha and Pritika Hingorani. We would like to thank Zehra Kazmi for reviewing the paper and providing editorial support. We are indebted to Omdiyar Network, without whom this work would not have been possible.

Suggested Citation

Pachisia, H.V., Bhansali, A., and Parasa, R. (November 2022). "The state of open data in urban India", Working Paper 01. Mumbai: Artha Global

The State of Open Data in Urban India

by Harsh Vardhan Pachisia, Anushka Bhansali, and Rajeswari Parasa[1]

July 2022

[1] Harsh Vardhan Pachisia was former Senior Associate at Artha Global ; Anushka Bhansali and Rajeswari Parasa were former Associates at Artha Global.

Abstract

Open data for cities, or urban open data, can build trust, promote innovation, and enable collaboration between municipal corporations, city residents and private companies. It can inform the economic and spatial planning of cities and support their sustainable development, acting as a crucial input for emerging tech-enabled governance tools. An analysis of the existing urban open data ecosystem is the first step towards this goal. This study fills that gap by investigating the state of open data in urban India through a people and data lens. It identifies the different 'people' or stakeholders in the ecosystem, and maps them according to their power and interest in developing urban open data. It then discusses three major 'data' aspects that need improvement: the lack of city open datastores or data platforms, data accessibility and the limited albeit emerging use of new technologies, such as GIS, in urban governance. Using case studies, the paper demonstrates how improving these data aspects can enhance decision-making in our cities. The paper offers recommendations on how the Indian urban open data ecosystem can be improved in the short and long term. This study is not exhaustive and is based on secondary research.

Introduction

Cities are home to more than 50% of the world's population while occupying only around 3% of the land (United Nations, n.d.). Migration to cities due to the economic opportunities they provide will continue, with close to 68% of the global population projected to reside in urban areas by 2050 (United Nations Department of Economic and Social Affairs, 2018).

The United Nations has recognised this transition, with Sustainable Development Goal 11 aiming to "make cities inclusive, safe, resilient and sustainable by providing accessible transportation, supporting economic linkages between peri-urban and urban areas and integrated policies" by 2030 (United Nations, n.d.). This ambitious target can only be achieved if policymakers make informed decisions based on data-driven insights into evolving urban systems. However, these insights are sorely lacking, especially in India, the world's second-biggest urban system, with 11% of the total global urban population living in its cities.

As Amitabh Kant, CEO of NITI Aayog writes, "Urbanisation will be the single biggest agent of growth in the next few decades in India"; the UN estimates that around 416 million people will be added as urban dwellers in India between 2018 and 2050, and the country will be more than 50% urban by 2050 (Kant, 2022). Moreover, beyond their burgeoning populace, Indian cities are the country's engines of economic growth, contributing close to 70% of its GDP (Mitra & Mehta, 2011).

Despite the importance of cities, their potential is unrealised due to poor urban governance, with issues of planning and management and poor implementation capacity. Indian cities are often sprawling and unorganised, with substandard infrastructure and outdated processes to manage them.

One of the key reasons for poor governance of Indian cities is the lack of organised open data systems, a fact laid bare during the country's response to COVID-19. The crisis devastated urban areas, revealing fundamental flaws in how city officials collected, disseminated, and used data in policy decisions. The COVID-19 crisis revealed how information gaps across sectors such as transport, public health, and social welfare led to subpar service delivery within Indian cities. For instance, lack of disaggregated spatial data on slums obscured any insight into their vulnerability or specific emergency needs (Sachdeva & Patel, 2022).

Urban India requires a post-COVID-19 plan to build systems that collect, disseminate, and utilise data to improve service delivery. Developing such 'data-enabled cities', as coined by the Open Data Institute (ODI), will also improve policymakers' transparency and accountability, lead to better decision making, and ultimately, enable innovation in urban governance (Snaith et al, 2021). Through a mix of open data and transparency around urban planning and policymaking, officials would be able to build a wider ecosystem of skilled partners, overcoming internal capacity issues, as well as catalysing innovative solutions. For instance, start-up hubs, public-private partnerships, and data challenge programmes can lead to collaborations between various stakeholders for solving problems in Indian cities, wherein each of them has an incentive to participate in the project. Developing citizen engagement platforms such as vTaiwan and Decidim (Barcelona) would help planners communicate directly with residents most affected by their proposed urban plans. Moreover, many cities are now 'planning for openness' –Glasgow announced that its urban plans were based on the principles of open government (Glasgow City Council, n.d.); Helsiniki has based its city operating model on openness and transparency.

Building solid open data ecosystems should also help future-proof cities and form the backbone of emerging technologies in city governance. We seem to be rapidly approaching another industrial revolution in city governance with the advent of digital twin cities (DTC) (Deng et al, 2021). DTCs are essentially digital parallels which map physical entities. They have the potential to integrate existing technologies such as Geographic Information Systems (GIS), Internet of Things, and 5G, among others. Indian cities should aim to lead such a revolution, given that they will house a significant percentage of the global population – people whose quality of life must be improved while ensuring that their rights, such as privacy, are maintained.

Ultimately, policies directed at the growth of India's urban open data ecosystems to build smart cities and conduct data-driven governance require analyses of the existing landscape. This paper attempts to fill that gap by investigating the state of urban open data in India through a people and data lens. First, we analyse the different stakeholders in the ecosystem, in terms of their power and interest in the ecosystem. Next, we discuss three major data aspects of the ecosystem that need improvement– the lack of open city datastores, data accessibility issues, and the limited use of new technologies, such as GIS, in urban governance. Using case studies, we demonstrate how improving these three data aspects can enhance decision-making in our cities. Finally, we discuss the implications of our findings and present recommendations on how the urban open data ecosystem can be improved.

Background

Cities are a 'system of systems' with multiple subsystems such as transportation, retail, health, welfare, law & order, finance, water, etc (Johnson, 2012). These subsystems are deeply intertwined and keep evolving as the city grows. However, if one system is poorly run, it often leads to negative externalities for the development and smooth functioning of the entire city. To prevent this, all such systems require information to be collected, analysed, and disseminated.

Globally, organisations and cities produce nearly 1.7 million billion bytes of digital datasets per minute. That is the 'data' that is generated. Only 4% of this information is made available in the public domain (i.e. made into 'open data') (Rial, 2013). For Indian cities, this number is possibly even lower, since most data about India is collected at the national, state and district level. This hierarchy completely sidesteps cities, often leading to a lack of information on labour markets, transport, economic activity, and housing within cities. We simply don't know enough about our cities to make data-driven policies. As a result, urban governance suffers.

Sadly, this has always been the case. Cities found the deserving policy focus and grants allocation only after the launch of Jawaharlal Nehru National Urban Renewal Mission (JNNURM) in 2005. Until then, the focus was largely towards the rural (Jha, 2020). While urban development is constitutionally a state subject, the national government primarily sets the policy agenda and rolls out various schemes and initiatives for promoting urban infrastructure development (Praharaj & Bandyopadhyay, 2020). This extends to policies that affect open data released by cities such as the National Data Sharing and Accessibility Policy (NDSAP), 2012 as well as platforms, such as the Open Government Data Platform in 2014 which acted as a one-stop-shop for all open data published by government departments and ministries.

To its credit, the Ministry of Housing and Urban Affairs (MoHUA) has developed its open data portal titled SmartCities.data.gov.in for urban India (where cities can publish data) and has run initiatives such as DataSmart Cities aimed at building data capacity within municipal corporations. It also tried to communicate the benefits of publishing such data via events such as the Open Data Initiative Week (MoHUA, 2022).

However, such programs often turn into haphazard, compliance-oriented procedures with emphasis on the number of datasets collected rather than their quality or usefulness. For example, the SmartCity portal was launched in 2015, but has a relatively limited number of downloads (~100,000) of datasets. To correct this issue, MoHUA has recently launched initiatives such as the National Urban Digital Mission and the Indian Urban Data Exchange (IUDX). Both programs have a use-case centric approach, with data only being collected after identifying a service that would be improved through that data. Civil society organisations such as Tata Trusts have also created city-level open data platforms through collaborations with city governments to facilitate sharing of data between policymakers, industry, citizens, and academia.

The urban open data ecosystem has a variety of stakeholders with different motives, incentives, and power to influence its overall growth and encourage the collection, dissemination, analysis and use

of city data for improved governance. Hence, Indian cities require a structural shift in thinking about data and its use in evidence-based policymaking.

Research Questions

Open data studies in India, and globally have, by-and-large focused on evaluating national-level open government data platforms and policies. Similar studies at the city level are scarce, especially in the Indian context, despite the economic and human significance of urban development and smart cities. Our study contributes to bridging this gap in India by investigating the landscape of this ecosystem and answering two research questions:

1. Who are the 'people' (i.e., stakeholders) present in the urban open data ecosystem and what is their role and interest within the space?
2. What specific 'data' aspects, both from a governance and technical standpoint, need improvement in the urban open data ecosystem and how can such improvement enable evidence-based decision-making in cities?

It is important to note that this is not an exhaustive study of the urban open data ecosystem in India and is based on secondary research and a review of initiatives undertaken by ministries and stakeholders in this space. The limitations of the study are further elaborated at the end of the paper.

People | Stakeholder analysis

Before any improvements can be undertaken in the urban open data ecosystem, its prevailing structure needs to be understood. To that end, we analyse the 'people' in the space by applying the open government data stakeholder analysis framework built by Zapata and Heeks (2015) to the Indian urban data ecosystem.

Zapata and Heeks' framework looked at all the open government data stakeholders in Chile and their roles in the system. . Then, the framework maps all the stakeholders based on their power and interest in this space. For this study, we first identify stakeholders at the forefront of shaping the urban open data ecosystem in India. Next, like the Zapata and Heeks framework, we study each stakeholder's 'power' and 'interest' in developing the ecosystem. We define power as the ability to influence other stakeholders in the urban open data ecosystem and interest as willingness to publish and use data, and help the urban open data ecosystem develop further.

As we conducted the analysis, it became clear that a one-size-fits-all metric was not applicable as stakeholders have varied power and interests. Therefore, this analysis is based on secondary research, illustrative examples, and initiatives undertaken (outlined in Appendix A) to make city data publicly available or more accessible. We also highlight the absent stakeholders that we believe should be a part of the system, but are currently missing from the space.

The 'People' in the Ecosystem

The urban open data ecosystem has various stakeholders (Figure 1), with disparate degrees of power and interest (Figure 2). Their competing motives are often responsible for either improvements or setbacks in how the system evolves. Figure 2 maps the power and interest of stakeholders in improving and developing the urban open data ecosystem. Note that this mapping is based on our analysis/interpretation of open data initiatives and projects. We have also not included the absent stakeholders in Figure 2, since they are currently not part of the system

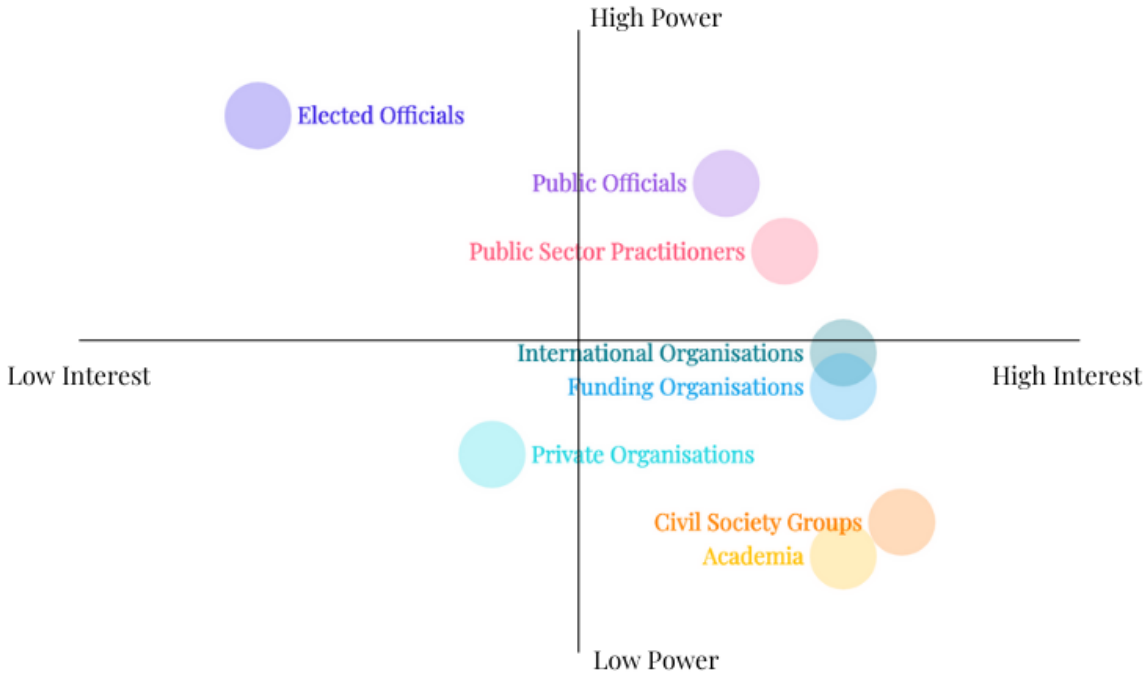
.

Figure 1: Stakeholders of Urban Open Data in India

Stakeholder	Definitions
Elected Officials	• Members of Parliament (MPs) • Members of Legislative Assembly (MLAs)
Public Officials	• Ministries • National Government Bodies • Secretaries
Public Sector Practitioners	• City Officials • Municipal Commissioners • IT officials at Municipal Corporations
Civil Society Groups	• Civil Society Organisations • Volunteer Driven Networks • Non-Government Organisations • Non-Profit Organisations
Academia	• Universities • Research Institutes
International Organisations	International Agencies and Companies
Funding Organisations	• Philanthropies • Trusts
Private Organisations	Private Agencies and Companies

Source:Authors' Estimates

Figure 2: Stakeholder Analysis of Urban Open Data in India



Source: Authors' estimates

Elected Officials

Elected Officials have a high level of power since they contribute to decision-making on the type of data to be collected, legislation around open data and the financial approval to build these systems.

Such power to set policy agenda often resides at the national level, despite urban development being a state subject in India (Praharaj & Bandyopadhyay, 2020). In recent times, elected officials have recognised the importance of "modernising" and nurturing Indian "megacities and their hinterlands so that they become current centres of economic growth" (Ministry of Finance, Budget 2022-23, para. 68). As a result, there have been substantial outlays on improving urban infrastructure in Budget 2021-22, and the way our cities are planned in Budget 2022-23 (Jha, 2022).

The government has signalled that it considers data an important component of evidence-based policy that furthers the larger urban agenda. Hardeep Singh Puri, Union Minister of Housing and Urban Development, stated that "data is becoming increasingly significant for improving city services" (Press Information Bureau, 2020).

While elected officials' interest in using data for city governance is rising, their appetite for making information publicly available seems to be limited. For instance, elected officials might not want to disseminate statistics that cast them in a poor light such as employment trends and potentially sway

voter opinion. On the other hand, open data might simply be a low priority for elected officials, who have largely left the evolution of the Smart Cities program (into DataSmart Cities, and National Urban Innovation Stack, among others) to public officials.

Public Officials

Public officials, especially at the national level, set the policy agenda for urban governance in India. They have been largely responsible for how data is collected, used and disseminated within Indian cities via a wide range of policies and programs (Appendix A).

Public officials at MoHUA oversee the creation and evolution of programs that determine the future of most urban India. These include initiatives such as Smart Cities, DataSmart Cities, India Urban Data Exchange, and National Urban Innovation Stack, among others. They can, hence, effect change at the city and municipal levels.

For example, at the "Smart Cities, Smart Urbanisation" conference for city officials in April 2022, a Digital Governance Pavilion was organised that included competitions on open data, AI/ML cases and the use of platforms for cities such as the India Urban Data Exchange (Times of India, 2022). MoHUA also organised a role-play, data game for all city officials to understand the inherent value of urban open data.

Such initiatives demonstrate the interest and intent of public officials to promote open data. Despite their limited capacity and stretched bandwidth, public officials remain the de facto central stakeholders in pushing the urban open data ecosystem forward. However, it should be noted that the transferability of public officials between ministries does pose a contingent risk on the continued development of the space and limits their power to a certain extent.

Public sector practitioners

The policies formulated by public officials are executed by public sector practitioners at all levels of government (national, state, and city). These practitioners are instrumental in implementing city-data policies and schemes. They oversee the collection, analysis, and dissemination of data used for policy implementation. In Surat and Pune, for example, city officials have taken advantage of national-level schemes such as Smart Cities and its associated programs, to create some of India's only city open datastores and pioneer data-driven governance. Using Pune as an example, we look at two key practitioners in implementing urban open data policies in the city.

	Role	**Initiative**
City Data Officer (CDO)	Formulates data collection and dissemination strategies at the city level under Smart Cities Mission.	Data collection and dissemination on the smartcities.data.gov.in and Pune Municipal Corporation (PMC) Open Datastore

IT Departments of Municipal Corporations	Maintaining the city datastore and thereby, responsible for the dissemination of data.	PMC Open Datastore The IT department of PMC maintains and hosts the open datastore with 559 datasets.

Figure 3: Illustrative examples of Public Sector Practitioners

However, Pune is the exception, not the norm. In most cities, public practitioners lack incentives, capacity and interest to significantly improve open data except certain self-motivated champions.

Academia

Universities and research institutions complement government efforts, from analysing gaps in the ecosystem to improving data accessibility, building city open datastores and demonstrating the value of open data in urban governance through analysis. Further, specific departments and professors assist governments (at the national, state and city level) in developing and implementing data-driven city governance programs. For example, IUDX is housed and assisted by professors from the Robert Bosch Centre at the Indian Institute of Science (IISc), Bangalore.

Academics are critical for better accountability, but also for improved innovation and knowledge creation. Therefore, they have a high interest in this space. At the same time, they cannot influence the space without buy-in from public officials or public sector practitioners. In Appendix A, we outline the role and initiatives taken by some of the universities and research institutions.

Civil Society Groups

Civil society organisations and data volunteer community stakeholders complement the efforts of research institutions and universities. At times, they also become substitutes for the state's lack of data capacity. This was underlined by how covid19india.org (a volunteer-driven effort) became the primary source of data on COVID-19 caseload for both government officials and researchers due to data gaps in traditional sources, with the portal actively used for making policies in real-time.

In general, the activities of civil society in this space can be bucketed into three categories. First, they improve the accessibility of publicly available government data by cleaning, collating, and disseminating them as machine-readable repositories of information. Second, they showcase the potential of using traditional and alternative data sources for policymaking via use cases and case studies. An example includes CivicDataLab, an organisation that partnered with the Oorvani Foundation to re-design the Open City Platform to make it more accessible to users and publishers of the platform. Third, such organisations create, and at times, operationalise data standards within the government by closely working with relevant stakeholders at various administrative levels. One example is Tata Trusts, which partnered with Tata Consultancy Services and worked with Pune Municipal Corporation (PMC) to onboard their first Chief Data Officers (CDO) (Tata Trusts, 2019

p.3) and held capacity building sessions for department officials that led to increased knowledge on data standards (Ibid. p.7).

Civil society organisations are key to promoting the usefulness of open data for improved decision-making, transparency and innovation. Despite their high interest, like academia, they possess low independent power to influence the space without buy-in from public officials or public sector practitioners. Some of the specific organisations involved in the space are outlined in Appendix A.

Private companies

Private companies in the urban open data ecosystem play three roles, with varied levels of interest. These are data users, data intermediaries, and data providers through data collaboratives (Gurin et al, 2019). First, a lot of small and medium enterprises (SMEs) benefit from using government data for building their business models which they cannot collect themselves due to inherent capacity constraints. For instance, out of the 500 global companies mapped by GovLab which utilise open government data, over 70% of them were SMEs and start-ups (Verhulst & Caplan, 2015). An example in India is Spageo, an engineering company that uses geospatial, vehicular and infrastructure data from the Government of India.

Next, as data intermediaries, companies often improve and collate government data for public use and derive insights that are useful for other businesses and entities. An example is IndiaStat, which requires users to pay a subscription fee to access cleaned and collated socio-economic government statistical data.

Third, a few private companies have become direct data providers in the urban open data ecosystem. These include private companies such as the Centre for Monitoring Indian Economy (CMIE) which provides a range of services – from primary data collection to analysis of government data. For instance, it is often used as a single source of truth for labour market/employment data for Indian cities (despite charging a fee for accessing its data) due to either a lack of such critical statistics provided by the government or poor dissemination.

Other global examples include Uber Movement and Google Mobility (a powerful source of information during COVID-19) that publicly release data on mobility for certain Indian cities based on large-scale data collection. Such large companies have also come under pressure from the Indian government to share their other internal datasets publicly in recent years.

The private sector seems to be split in terms of their interest in the development of the urban open data ecosystem. While start-ups have high interest as they rely on such data due to a lack of in-house capacity and would like it to develop, data intermediaries have limited interest since it might cannibalise parts of their business model. On the other hand, data providers have medium interest as they could potentially welcome government data that assists them in improving their services. Ultimately, the private sector has low power and medium interest in space.

International Organisations

International organisations are key in developing the urban data ecosystem, due to their dual role of being financiers as well as implementation partners. As financiers, they bring to life large scale city projects that improve data collection and usage. For example, the World Bank approved a $150 million partnership with the city of Chennai to make the city more resilient in the areas of water supply and sewerage, mobility, health, and solid waste management (World Bank, 2021). Across these areas, the partnership has emphasised the role of collecting, using, and disseminating data to improve service delivery.

International organisations also play an important role in advocating for and implementing data standards across cities globally. For example, the City Data for India Initiative by Tata Trusts in partnership with World Council for City Data (WCCD) helped Pune, Surat and Jamshedpur commit to building high calibre data in Indian cities. It did so by committing them to the ISO 37120 (Sustainable Development of Communities: Indicators for City Services and Quality of Life) certification standard (Tata Trusts, 2017). In 2018, five more cities, Ahmedabad, Bhubaneswar, Bhopal, Chennai and Vijayawada joined this initiative (Tata Trusts, 2018).

Since they offer both funding and implementation support, international organisations hold medium power and high interest in the development of the urban open data ecosystem. They understand the landscape and have the necessary resources but need to adhere to the larger policy agenda set out by the government and public officials.

Funding donors

Funding donors occupy both direct and indirect roles in the urban open data ecosystem. At times, they directly fund other stakeholders such as civil society, academia, and private companies working to develop the space. For instance, the E-Gov Foundation, a pioneer in developing data systems in Indian cities through various platforms on municipal services, is funded by organisations such as the Bill and Melinda Gates Foundation, Nilekani Philanthropies, Omidyar Network India, and Tata Trusts.

Donors also indirectly impact the general data ecosystem by mandating grantees to carry out evidence-based project monitoring and reporting (Chattapadhyay, 2014). This creates a data culture within civil society organisations and non-government organisations funded by donors. However, this has a limited impact on the urban open data ecosystem because the donors may not mandate publishing the data collected as part of their projects.

Donors can display a high interest in expanding the space but possess medium power in the ecosystem since their success is dependent on grantees implementing and institutionalising initiatives in the long term.

Absent stakeholders

A notable absentee in the Indian urban open data ecosystem is the individual citizen-user (i.e., the residents of Indian cities). We could not find instances where private citizens have engaged or

contributed to developing urban open data. Most commentary and interest in the space stems from volunteer groups and nonprofits active in the space, private companies, and academia. This might change in the future as Indian cities become more tech-enabled and awareness about data governance develops further in the citizenry. At present, however, they remain silent stakeholders despite the indirect impact urban open data has on their lives through various public services.

In the larger stakeholder group of elected officials, elected representatives in the urban local bodies, such as the ward councillors, are also missing from the urban open data landscape. They are actively involved in civic issues such as road laying, disposal of waste, involvement in matters of town planning, and so on and are responsible for the welfare of the people they represent. Hence they are uniquely positioned to enable data collection and dissemination activities at a granular level in the city. However, currently, the data policies and initiatives at the city level do not seem to involve them in a suitable capacity.

While citizen interest and the role of elected representatives in this space are missing across urban India, there is significant variation across cities. For instance, while public sector practitioners are keenly involved in developing urban data in Pune via their open datastore, this is not the case for most Indian cities.

Ultimately, the urban open data ecosystem is still largely in its infancy and its development will take time. The stakeholders involved in the space must collaborate and work together to build and expand open data, and improve the 'data' aspects that are outlined in the next section.

Data | Governance and Accessibility

There are a variety of 'data' aspects, both from governance and technical perspectives, that need improvement for the development of the urban open data space. We analyse three such aspects in this section. These include the lack of city open datastores, the importance of data accessibility during emergencies, and the potential of using new technologies, such as GIS, for making data-driven policies. Through case studies of Indian cities, we demonstrate how improving these three aspects can boost evidence-based decision-making.

The lack of City Open datastores | Pune is the outlier, not the norm

A city open datastore is needed to "provide a common, reliable evidence base to inform city decision-making and improve sharing with all levels of government and non-government organisations" (Global Smart Cities Alliance, n.d.). While they are inherently useful for policymakers, such stores also provide key information to different stakeholders such as civil society, academics and entrepreneurs which they can use to offer solutions to long-standing urban problems. For example, the open datastores of New York and London are often regarded as vital enablers of innovation in urban planning and public services.

At present, such stores are scarce in India despite initiatives taken by MoHUA to make municipal corporations release data at the city level. First, as mentioned earlier, the smartcities.data.gov.in portal that hosts data from 100 cities runs the risk of becoming a compliance exercise rather than a useful datastore. More recently, the IUDX, which hosts data for 17 cities so far, displays promise due to a use-case centric approach rather than a quantity of datasets one. However, the portal is still in its infancy and data for public consumption is extremely limited. Further, while some cities such as Solapur have implemented city-data policies (Solapur Municipal Corporation, 2019), the collection and collation of these datasets into an integrated city datastore is often missing.

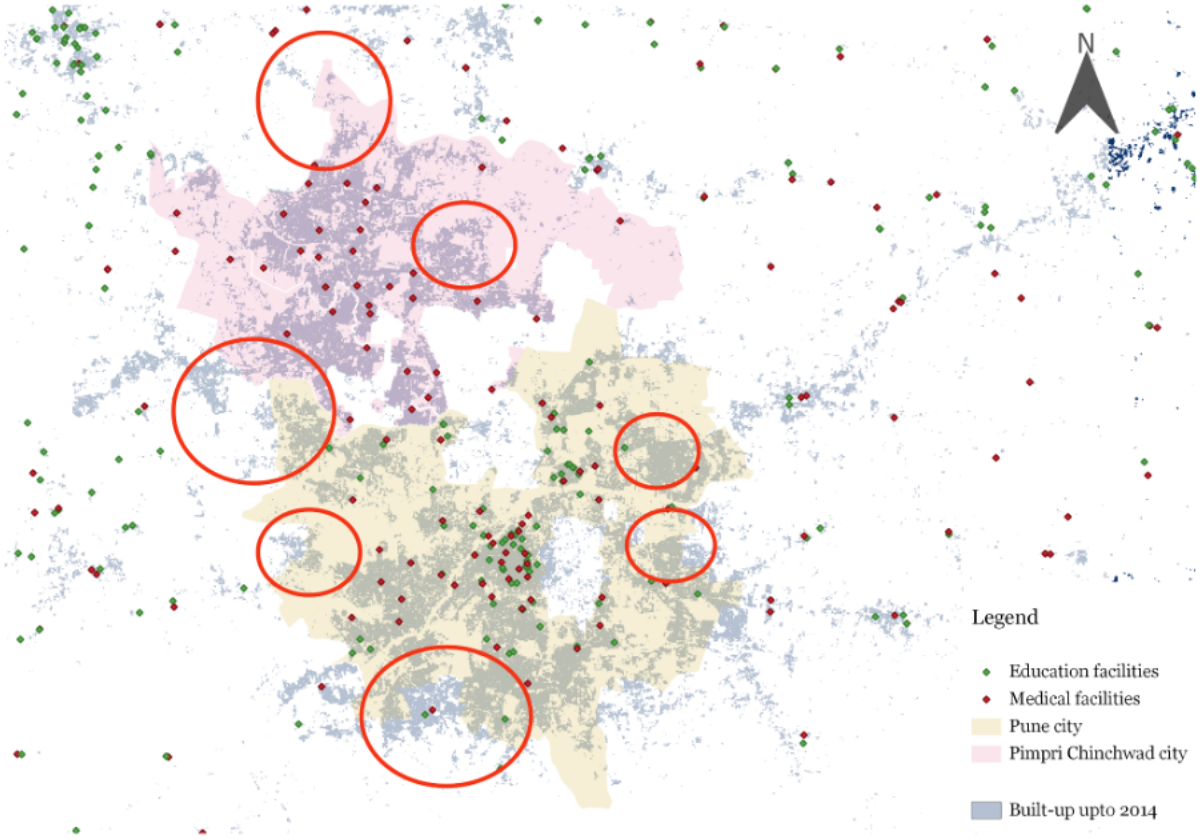
The lack of such stores could be a result of capacity constraints, or a lack of incentives and bandwidth of overworked city officials. At the same time, certain Indian cities such as Pune and Surat have recognised the potential of developing open datastores, becoming leaders in the field. We use Pune as a case study to demonstrate the potential usefulness of the data it collects and publishes on its datastore.

Case Study: Pune's Open datastore and measuring accessibility in peri-urban areas

Pune is the first city to host an open data platform or 'datastore' with 559 datasets hosted across 27 departments, maintained wholly by the Pune Municipal Corporation as of April 2022. Through information collected from various city departments such as public health, transport, education, etc., Pune has been able to establish points of interest (POIs). That is, the latitude and longitude of each healthcare and education facility (among others), has been collected and disseminated on the Pune Open datastore.

As Pune continues to develop, such POI information is extremely useful in analysing the accessibility of such facilities on the outskirts of the city. In this case study, we use the POI datasets from Pune's open datastore, the smartcities.gov.in platform, and the PMGSY GeoSadak portal (an online Geospatial Transaction System for Pradhan Mantri Gram Sadak Yojana program) to study the distribution of such facilities in the urban areas of Pune, Pimpri Chinchwad, and neighbouring rural areas. We overlay the medical and educational facilities and the built-up layer via Global Human Settlement Layer data in these areas (Global Human Settlement, n.d.) in Figure 4.

Figure 4: Map highlighting some of the probable areas of service delivery blackspots in Pune and neighbouring areas



Source: Authors' estimates

We observe that the density of facilities decreases towards municipal corporation boundaries and becomes much sparser in rural areas, with an inverse correlation with the built-up area (i.e., more development equals the higher density of facilities). However, there are also regions lying immediately on either side of the city boundary that have considerable built-up but barely any facilities. These could potentially be densely populated urban areas lacking sufficient facilities. Such areas can be the service delivery 'blackspots', requiring attention from urban planners. Further analysis factoring in parameters such as land use (residential, commercial, etc) and demographics should be conducted to identify the specific locations of the blackspots to improve service delivery there.

The illustrative example above highlights that as India (not just Pune) continues to rapidly urbanise, managing the growth of our cities through better urban planning and improving access to services will be key. To that end, developing well-managed, integrated open datastores that provide information on areas such as locations of facilities are needed to form the evidence base of new policies and private sector innovation.

It should be noted that while Pune is a pioneer when it comes to open datastores in India, it lags far behind global standards when compared to the London open datastore. For instance, the Pune store lags in the number of urban datasets available, with "the difference especially staggering in the areas of demographics, environment, jobs and employment and health data availability" (Praharaj and Bandyopadhyay, 2020) (details on challenges faced while accessing the data are outlined in Appendix B). Moreover, Pune shies away from the publication of sensitive data such as crime, jobs growth, and road accidents as well when compared to London (Ibid.).

Ultimately, while there is still a long way to go, Pune has challenged the notion that Indian cities lack the technical know-how and capacity to collect, collate and communicate data in an organised manner by becoming the first open datastore managed fully by the municipal corporation. The same cannot be said of other Indian cities, a trend that needs to change.

Making city data accessible and enhancing dissemination | Case of BMC COVID-19 data

The COVID-19 pandemic highlighted a systemic problem in how Indian cities use data for policy making (Pachisia & Gutta, 2021). During the crisis, it was critical to make quick, well-informed decisions as the situation on the ground was rapidly changing. However, due to a general lack of systematic data systems within cities, information on key parameters such as caseload, testing, vaccination, deaths, etc. was not available. As a result, local government authorities, researchers and policymakers found it difficult to plan, manage and make forward-looking decisions due to the lack of data.

Even when city governments were able to collect data, it was not communicated properly. For instance, critical information on caseload and bed capacity was released haphazardly on Twitter handles in the form of images or scanned PDFs, making it difficult to analyse and use the numbers as a basis for future decision making.

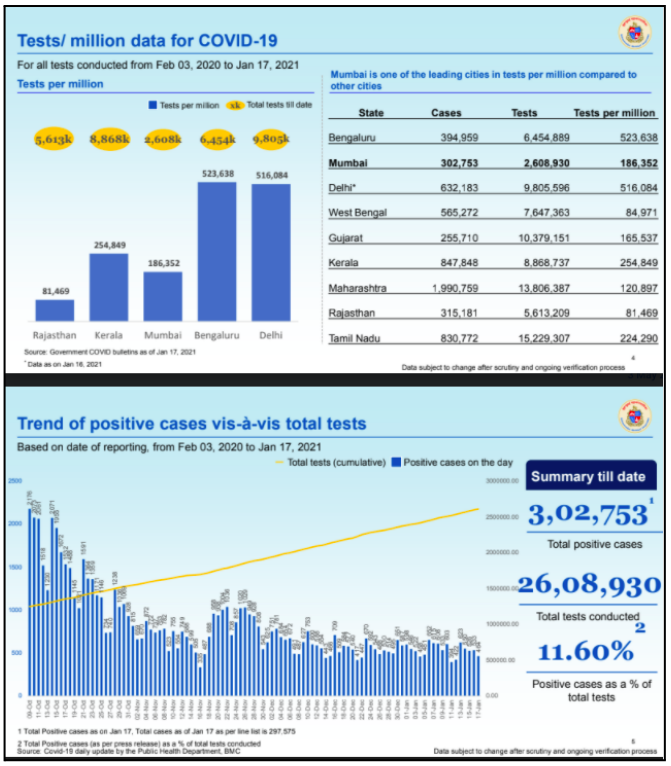
During the crisis, civil society stepped in as a substitute for the state's lack of data capacity. A prime example was the covid19india portal, as mentioned earlier, a volunteer-driven initiative that collated information from various official ministry websites, social media accounts, bulletins, etc on key indicators at a district level. This volunteer-driven initiative became the single source of truth for COVID-19 data in the country with about 1460 mentions on Google Scholar as of 13 April 2022.

While covid19india largely focused on the district level, municipal corporations in various cities were also collecting information but not disseminating it effectively, leading to civil society organisations stepping in to improve the accessibility of such data. The Brihanmumbai Municipal Corporation (BMC), for example, despite excellent data collection, fell short in providing information in a publicly accessible manner. The authors were part of a team at the IDFC Institute that created a public facing database (IDFC Institute, 2021) and data stories based on the daily bulletin put out by the BMC. In this paper, the authors highlight the lessons and insights from this larger effort and present a case study on the importance of data accessibility, not just collection.

Case study: Improving the accessibility of Mumbai's COVID-19 data

The megacity of Mumbai, home to over 20 million residents, became a hotbed of COVID-19 cases in April 2021. To keep track of the crisis, the BMC published daily, detailed PDF bulletins on testing, bed capacity, caseload, vaccination, containment zones, etc. intending to share accurate and timely information regarding the pandemic (Figure 5). This was a step in the right direction for making faster evidence-based policy decisions during the pandemic. However, there were major gaps in dissemination and accessibility which almost negated their efforts, leading us to step in and develop a public-facing database that tracks the nearly two-year journey with COVID-19 from June 2020-September 2021. The database tracked daily metrics like testing and contact tracing, quarantine and containment measures, mortality trends, occupancy in hospitals, and vaccination.

Figure 5: Snapshot of a daily Bulletin published by BMC



Source: BMC

Developing this database was no cakewalk. The first issue with the BMC's dissemination was that the government body seemed to have made an internal decision not to maintain an archive with a historical record of all reports released. BMC would publish the new detailed bulletin each day replacing the previous bulletin. Therefore, only short-term decisions could be taken based on the initiative and temporal analysis and long-term research was not possible.

Second, while PDFs (the format of these reports) are a useful solution for sharing information on the web, extracting data from such files is a labour-intensive effort. This process became more of a challenge when the format of the report changed from PDFs to image files for a month, increasing the difficulty. Ultimately, **the creation of this database would not have been needed, had the**

data been published by the BMC in the form of archived spreadsheets instead of PDFs or images.

Third, quality checks and procedures had to be initiated to ensure that the data was available in machine-readable and usable formats. We had to initially create a repository of all daily bulletins and set up an automation for downloading the new report daily. Next, given that most of the PDFs were unreadable, we had to manually enter the data from bulletins to create a database of set indicators for each day. We conducted random quality checks by choosing indicators at random and re-checking to ensure the quality of data. After repeating this process for each day in the dataset, the final database was in machine-readable formats (Figure 6).

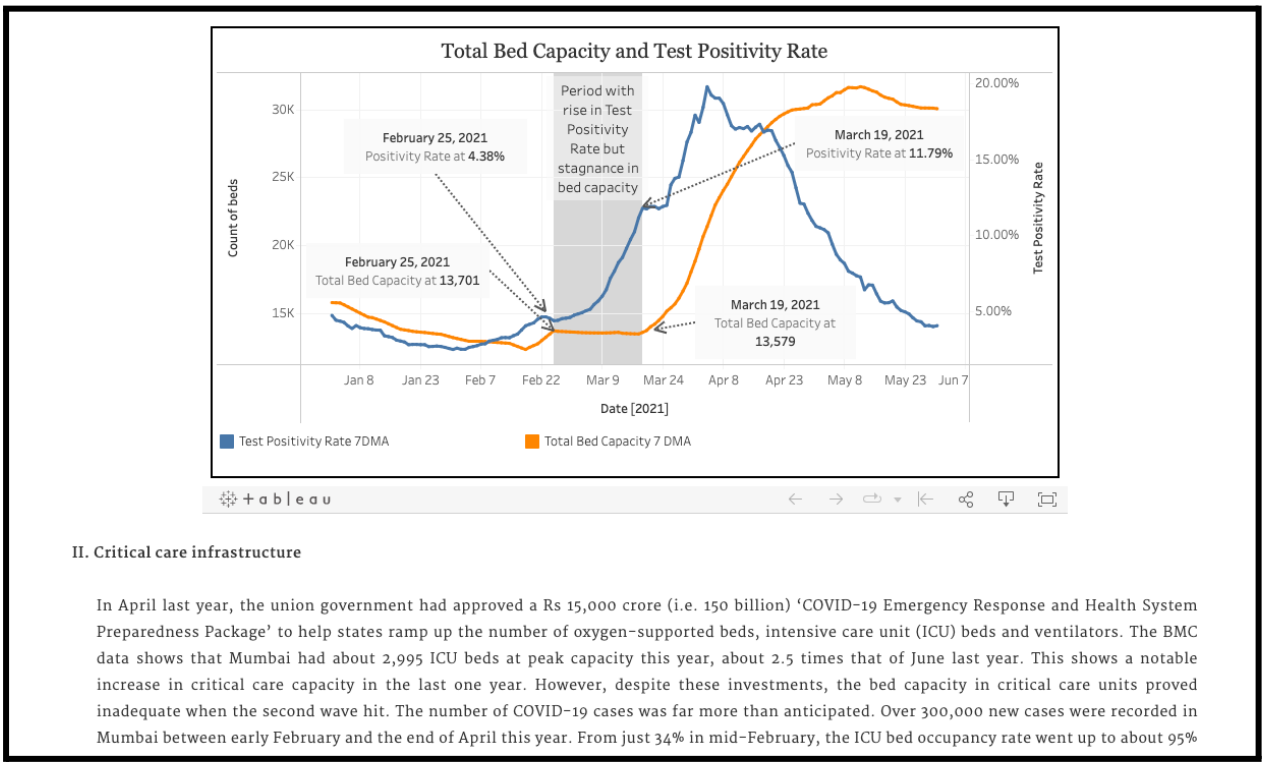
Figure 6: Sample indicators within the database

Date	Total_Bed_Capacity	Total_Beds_Occupancy	DCH & DCHC_ Bed_Capacity	DCH & DCHC_ Beds_Occupancy
2020 – 2021	11k – 32k	2.0k – 22k	6.0k – 23k	2.0k – 18k
October 21, 2020	18,477	9,870	15,107	8,584
October 22, 2020	18,477	9,673	15,107	8,396
October 23, 2020	18,417	9,492	15,047	8,249
October 24, 2020	18,476	9,282	15,106	8,101
October 25, 2020	18,476	9,089	15,106	7,911
October 26, 2020	18,476	8,777	15,106	7,680
October 27, 2020	18,476	8,553	15,106	7,482
October 28, 2020	18,476	8,372	15,106	7,328
October 29, 2020	18,476	8,191	15,106	7,167
October 30, 2020	18,315	8,060	14,945	7,042
October 31, 2020	18,362	7,817	14,992	6,841
November 1, 2020	18,362	7,667	14,992	6,765

Source: (IDFC Institute, 2021

Once the gap between data availability and accessibility was bridged, we analysed the data to provide key indicators and chart out Mumbai's journey through the pandemic. For instance, we observed Mumbai recorded the greatest number of deaths in June 2020, during the first wave, whereas its caseload was the highest in April 2021, during the second major wave. This threw light on the case management in both the waves and the importance of ramping up testing capacity (Figures 7 and 8). This offered an overview of important indicators and acted as a foundation for evidence-based policymaking while also providing other stakeholders such as academics and concerned residents with insights into Mumbai's journey through COVID-19.

Figure 7: Snapshot of a Blogpost by IDFC Institute



II. Critical care infrastructure

In April last year, the union government had approved a Rs 15,000 crore (i.e. 150 billion) 'COVID-19 Emergency Response and Health System Preparedness Package' to help states ramp up the number of oxygen-supported beds, intensive care unit (ICU) beds and ventilators. The BMC data shows that Mumbai had about 2,995 ICU beds at peak capacity this year, about 2.5 times that of June last year. This shows a notable increase in critical care capacity in the last one year. However, despite these investments, the bed capacity in critical care units proved inadequate when the second wave hit. The number of COVID-19 cases was far more than anticipated. Over 300,000 new cases were recorded in Mumbai between early February and the end of April this year. From just 34% in mid-February, the ICU bed occupancy rate went up to about 95%

Source: (Kwatra & Parasa, 2021)

Figure 8: Snapshot of the COVID-19 data story

*Note: It is important to note that there are studies by that estimate higher number of deaths in Mumbai during the first two years of the pandemic.

Source: (Bhansali et al, 2022)

This case study of Mumbai highlights two main points when it comes to data accessibility and the lack thereof within urban India. First, while it was a credible effort by Mumbai's city officials to provide such a detailed daily report on COVID-19, the hurdles we faced highlighted the government's inability to communicate critical information in an accessible manner. Second, our efforts (among other similar civil society efforts during COVID-19) underline the importance of publishing accessible and reusable data in real-time for cities, especially in an emergency. We find that at present, when cities do publish data publicly, either through a datastore or through the ULB's official website, the data cannot be easily used for temporal analysis or making long-term decisions.

To this end, the state needs to recognise its position in the data pipeline as being that of not just a data generator/collector but also a data disseminator. While civil society actors would always play a role in analysing and providing insights (i.e., complementing the state's efforts), they are not a sustainable substitute for the state's lack of data communication capabilities.

Ultimately, when city data is made publicly available, accessible, and well disseminated, it can be used as input for emerging technologies to assess the impact of large-scale projects and inform public investments. We showcase this in the following section.

Using GIS to inform decision-making | Case of Metro-3 and last mile in Mumbai

In recent years, India has started using emerging technologies for surveying and mapping, building information modelling (BIM) technology, IoT, etc. The most common has been the use of Geographic Information System or GIS, which is a very small subset of the technologies that can be used. Apart from GIS, unmanned aerial vehicles (also known as drones) have been used by state governments such as Andhra Pradesh and Chandigarh for detailed planning and governance. (PWC, 2020).

Collecting and disseminating machine-readable data about our physical surroundings enables use of such technologies. They have already been utilised around the world for improving urban planning and making faster, well-informed decisions on public service delivery (Esri India, n.d). For instance, Transport for London (TfL) provides data from various sensors to manage the uneven distribution of passengers in various stations (Open Data Initiative, 2016). It has a unified API in its open datastore that allows other developers to use its real-time data as a base on which they layer their unique services (Transport for London, n.d.). As a result, over 600 applications use the open data shared by TfL to improve service delivery.

The active uptake of technologies such as GIS that help urban planning requires the collection of data points to provide insights for policy decisions. If such information is inputted, it can also aid in assessing the impact and ROI of public investments. We provide a case study for the same below, using the Metro-3 line being built in Mumbai as an illustrative example.

Case study: The Metro line 3's impact on last mile connectivity in Mumbai

51% of the total trips in Mumbai city are made on foot (MCGM, 2016). Most public transport journeys (60%) also start and end as walking trips. Due to the large modal share of walking trips, the failure of metro projects to attain their projected ridership is often attributed to the lack of first-and last-mile solutions to desired destinations with metro stations (Bhattacharjee et al, 2020).

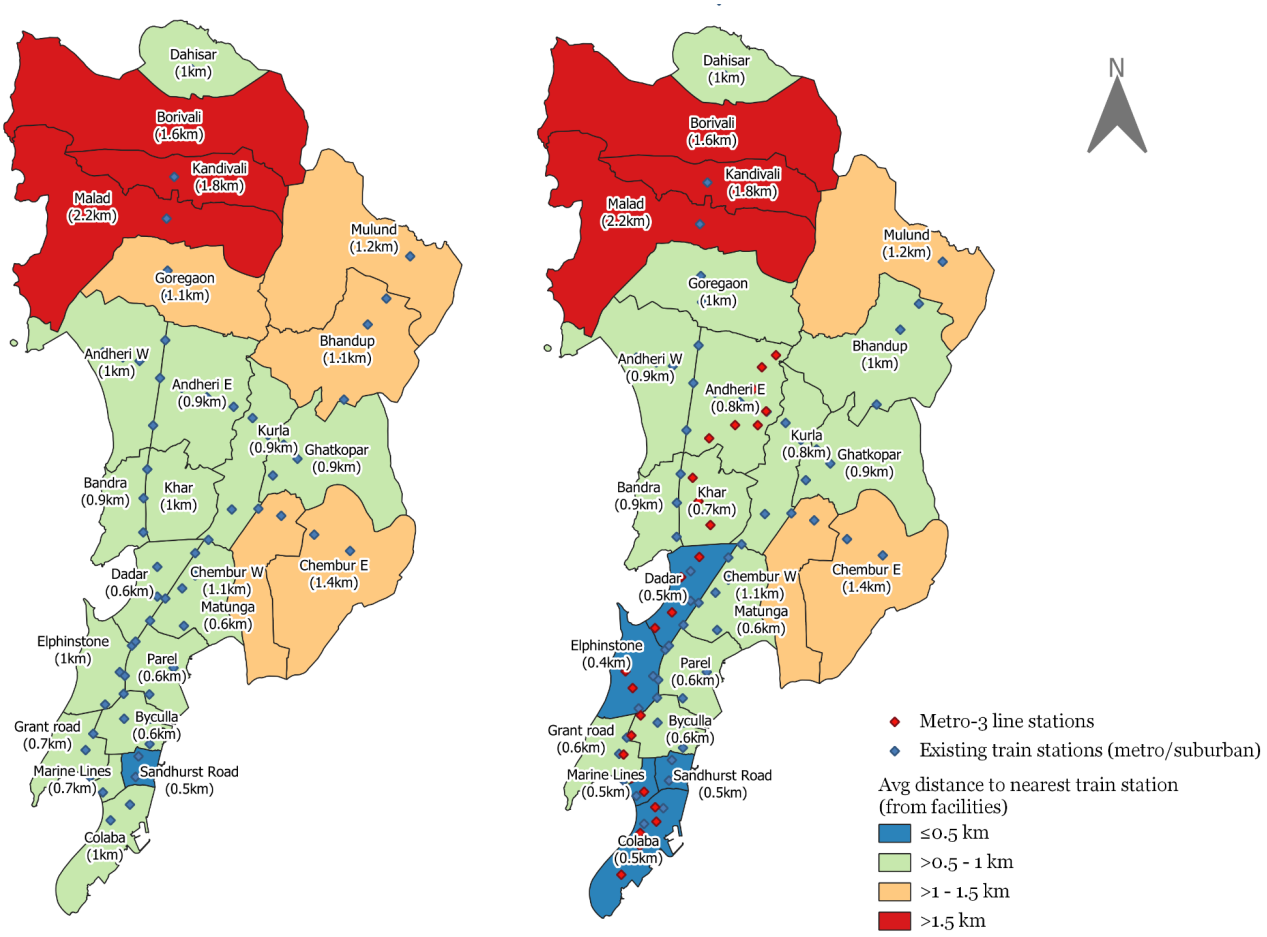
Mumbai is currently in the process of building out a new train line (Metro Line 3) that is meant to better link South Mumbai with the rest of the city. Given the problem of last-mile connectivity, it is a useful, live example to study the potency of using GIS to layer a variety of disparate urban data sources and showcase the actual impact that the new line has on solving the problem of door-to-door connectivity in the city.

We look at Mumbai's last mile connectivity by measuring the average distance between schools and hospitals and the nearest operational train station (suburban or metro) for each ward (detailed steps are given in Appendix B). We then measure the change in such distances following the under-construction Metro 3.

To achieve this aim, we first collated data and maps on POIs (schools, bus and train stops, medical facilities) and routes from a variety of sources such as the BMC (MCGM, n.d.),OpenStreetMap (a Wikipedia for mapping), Google Maps, Mumbai Metro Rail Corporation Ltd. (MMRCL, n.d.), etc. Given the lack of a Mumbai city open datastore, such collation from different stakeholders (public officials, private sector, civil society) cannot be avoided. Next, we overlaid all these datasets into QGIS, a software that enables us to analyse spatial data.

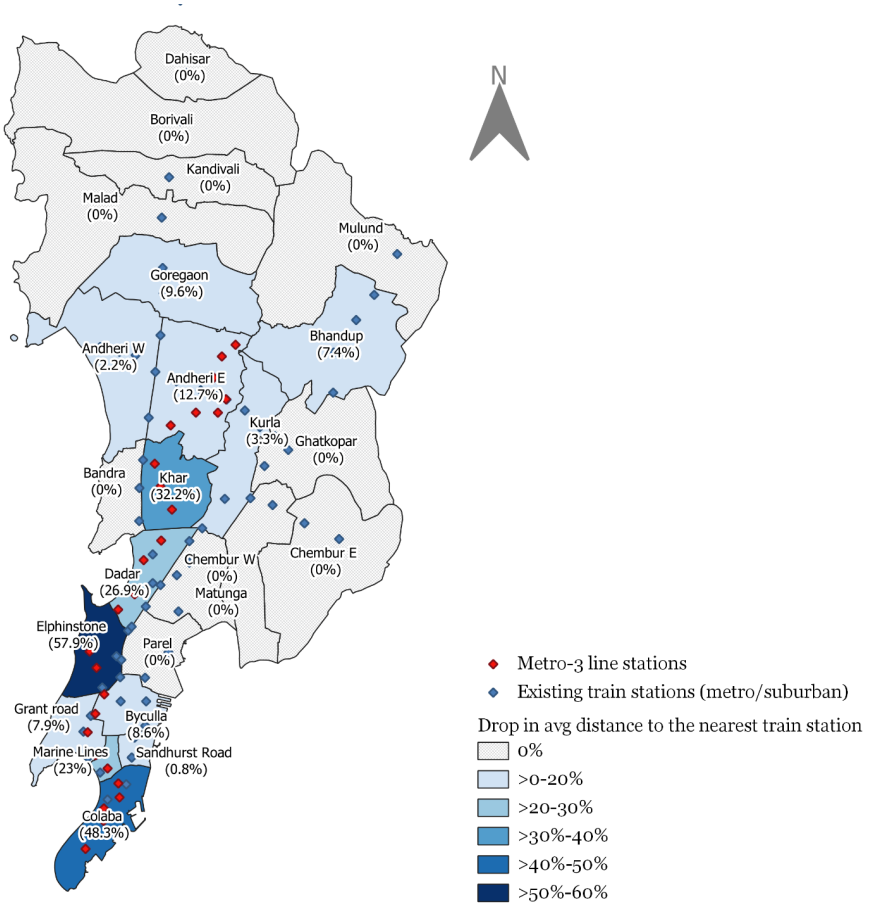
From our analysis, we find that both the schools and the medical facilities are at an average of ~1 km from the nearest train station in the city. With the proposed Metro 3 line, this distance drops by 6% for the schools and 9% for the medical facilities. Figure 9a shows the average 'last mile' or the average distance from the schools and hospitals to the nearest existing train station aggregated for each ward, before and after the Metro 3 project becomes operational. Figure 9b shows the percentage drop in the last mile that the project will cause.

Figure 9a: Analysis of the average distance from the schools and hospitals to the nearest existing train station



Source: Authors' estimates

Figure 9b: Analysis of the average distance from the schools and hospitals to the nearest existing train station



Source: Authors' estimates

We find that the largest drop in the average distance from a school or a hospital to the nearest train station will be in the southern parts of the city where the distance was already low, approximately 1km, before the Metro line 3 (detailed analysis of wards is given in Appendix B). However, the wards of Mumbai where the distance to the nearest station was the highest seem to be unaffected by the new line. Hence, while the metro seems to have generally improved connectivity, its limited impact on already disconnected places should be analysed further as a point of concern.

Similar spatial analysis can be performed to understand the spatially varying impact of Metro line 3 on the connectivity of people to their job locations by measuring the last mile to workplaces. However, due to the lack of granular open data on workplace locations, the present study limits the analysis to schools and hospitals.

Similar investigations for other planned and under construction metro lines in the city can help policymakers rethink investments, prioritise construction or operations among the multiple project lines, and seek public opinion. This limited example highlights the potential insights that using open

data (by collating and overlaying these disparate datasets together) and emerging technologies (like GIS) can have in planning, analysing, and assessing city projects and policies. Moreover, it showcases the importance that urban open data will play as data from disparate sources is required to leverage emerging technologies.

We do recognise the data privacy concerns that are inherent in the use of such technologies. While beyond the scope of this paper, we emphasise the importance of collecting only data points that are non-personal and non-human when using these technologies. The data privacy aspects of using such technologies will evolve depending on the implementation of the proposed regulations, such as the Data Protection Bill in India.

In general, if Indian cities can build capacity and buy-in for collecting and disseminating datasets properly (regardless of the use of new technologies), major improvements can be undertaken in how public services are provided, our cities are planned, and private innovation based on such data.

Discussion and Policy Recommendations

Before diving deep into each recommendation, we believe that the most pressing need to develop the urban open data ecosystem is to fundamentally change how data is collected in India. At present, data is mainly collected at the national, state and district level. This hierarchy completely sidesteps cities, often leading to a lack of concerted efforts in building and implementing standards for data collection, usage, dissemination and quality evaluation for city data. This results in lack of information on labour markets, transport, economic activity, and housing within cities. We simply don't know enough about our cities to make data-driven policies. As a result, urban governance suffers. Hence, statistical processes need to be built that ensure that information about Indian cities is collected, collated, and disseminated, especially given the rapid urbanisation taking place across the country and the economic importance of cities.

People | Align incentives, bring in absent stakeholders and engage users to create value

In this section, we recommend ways in which the key stakeholders can improve the urban open data ecosystem. An article by Global Open Data notes that implementation of an open data policy must be 'supported by capacity building, leadership, clear communication and a mixture of incentives to make the process of cultural transformation possible' (Smith et al, 2017). It is pertinent to build and align incentives for various stakeholders, including the absent stakeholders and engage current users of open data in the ecosystem to ultimately enhance their participation.

Communicate open data benefits to locally elected officials and define their roles in the ecosystem

A 2019 paper by Banerjee et al. found evidence from Delhi that public disclosures improve ward councillor spending and, in turn, their chances of re-election, thereby incentivising them to perform better. Granular open city data can create healthy competition among councillors to deliver better services to citizens and improve their chances of getting a party ticket. We recommend that such

information be communicated to elected representatives to showcase how useful promoting open data practices can be to their stay in office. Further, to actively involve them in the open data ecosystem, it's pertinent to identify and define their roles in the urban data policies and implementation guidelines, such as the City Data Policy (MoHUA, n.d.) under the DataSmart Cities Strategy, through a consultative process.

Recognise and leverage open data as a tool to promote innovation and economic growth

On the other hand, public officials should note that strengthening policies on city open data will not only improve trust and meaningful dialogue among stakeholders but also promote innovation and create economic opportunities (Open Data Charter, 2018). Consider the example of Transport for London (TfL). By making available free, accurate and real-time transit data such as timetables, routes, service status and disruption, TfL generated annual economic benefits and savings of up to £130m a year (Deloitte, 2017). As of October 2017, TfL's real-time data powered over 600 apps that were used by over 42% of Londoners and also created job opportunities (Ibid.).

Assess and address the needs of data officers across departments and build capacity

At the city level, roles such as city data officers are currently seen as additional responsibilities assigned to an official who is possibly already overburdened with other responsibilities. This must change, perhaps through an assessment of the bottlenecks of the data officers and their teams, suitable division of responsibilities and finally suitable capacity-building and training of city officials across departments. To overcome the inherent issue of retaining talent in the public sector, lock-in periods following skills-based training and better reward systems for data officers in cities should be explored.

Inculcate data-based decision making at the city level

MoHUA and NITI Aayog can also play a role in encouraging city officials to use open data in their day-to-day decision making. For example, the UK Government's open data experience shows that policymakers are one of the largest beneficiaries of open data (Pollock & Rogers, n.d.). In one instance, a senior UK government official used departments' open spending data at the transaction level to identify duplication in buying expensive market reports across departments (Ibid.). This insight saved the government £4 million from spending on the purchase of these reports. When officials realise such benefits of using open data, they would be more willing to open their datasets.

Encourage and incentivise non-government actors to collect and disseminate data

Funding donors also have an important role to play in enabling an open data culture among the city governments, non-government and civil society organisations. They should budget for, encourage and incentivise grantees on allocating a part of the funds to building related data systems and making them openly accessible. This can also reduce the duplication of similar data collection efforts by other organisations involved in related activities.

In the private sector, businesses benefit in multiple ways by opening access to their data (D'Addario, 2020) besides directly contributing to evidence-based planning and policy research. These benefits include improving market reach, making supply chains efficient, solving sectoral challenges, benchmarking processes and efficiencies, open innovations, building trust and so on (Ibid.). Governments can play a major role in enabling data-based collaborations and partnerships such as open innovation. For example, the European Union-funded Data Pitch initiative connects corporates and public sector organisations that serve as data providers, with start-ups and SMEs that work with data. SMEs are provided financial and advisory support to develop solutions to the problem statements put forth by data provisioning businesses and organisations. However, while enabling such collaborations, challenges observed globally (Susha et al., 2019) will have to be considered and assessed for local context. These include, issues of data ownership, lack of clear ethical guidelines and challenges in matching of data with a suitable problem (Ibid.).

Bring in absent stakeholders and engage users

Finally, a lot more needs to be done to bring in the biggest absent stakeholder, the city residents themselves, who ultimately face the consequences of a lack of open data. There has to be an increased focus on educating citizens about various kinds of open data and providing meaningful access to this data. NYC Open Data 2020 provides a useful reference here. It notes that resources accompanying a dataset, including a simple background of the data, tools such as maps and extensive documentation make the interpretation of data easier.

It is important to recognise open data as a potential collaboration tool to engage with citizens to co-invent solutions to policy challenges. This also furthers participatory governance. To its credit, MoHUA has recently organised events such as Open Data Week and Urban Geospatial Data Stories Challenge in order to build awareness around open data and engage citizens on proposing solutions to defined problems. Recently, the municipal corporation of New Delhi also held public consultations for providing feedback and suggestions on its proposed master plan, a policy that impacts its residents directly. While such efforts are welcome, a concerted effort needs to be undertaken in making these events and platforms sustainable solutions and collaborations. Further, students as citizen users should be actively encouraged to join the ecosystem and be involved from the start of their careers since they will experience the largest impact of India's urban transition. Such engagements would benefit the overall ecosystem by not only creating value out of the data but also informing the data practices of the government stakeholders.

Data | Improve dissemination of information and use of technologies

Create city open datastores, rationalise and converge efforts

Each municipal corporation should publish city data on its own platform, a datastore, or a regional/national platform based on the autonomy and financial resources available to the city's local government. The datastores should host spatial and temporal data across city sectors. The smaller

cities may benefit from publishing data through a regional or a national platform (Global Smart Cities Alliance, n.d.). In such cases, smartcities.data.gov.in or IUDX can be considered as potential platforms instead of hosting their own datastore. They should analyse Pune's example as a case study for understanding how it overcame capacity constraints in building its open datastore. Cities can also refer to NDSAP or the model policy created by Smart Cities Alliance for strengthening the open data policies at the city level (Ibid).

We also recommend that different city data platforms should make discoverability easier for users, reduce compliance overload, and duplication of efforts for data officers and contributors. For instance, the National Data and Analytics Portal by NITI Aayog launched in 2022 duplicates MeitY's efforts to maintain an open data portal in India. This is particularly important for smaller cities with smaller budgets. For example, Pune hosts city data on all three platforms – its datastore, smartcities.data.gov.in, and IUDX. All three different platforms demonstrate overlapping objectives of publishing user-friendly data for research and innovation. Such efforts should be streamlined.

Recognise municipal corporations' role as data generators and disseminators in the data pipeline

In the current ecosystem, non-government stakeholders, especially civil society groups and researchers end up substituting the role of government stakeholders in making data accessible (Pachisia & Ganapathy, 2021). We saw the importance of data accessibility during COVID-19. As a lesson from the pandemic, state governments need to formulate long-term data policies that institutionalise their role as data generators and communicators. While doing so, they should also align on standards for platforms, data, metadata and quality evaluation across cities and platforms to allow for comparisons and benchmarking. At the same time, these policies need to ensure they are consistent with data governance architectures like the Data Protection Bill that may change the landscape of the Indian data economy.

Build a culture of GIS-enabled data-driven decision making

Using GIS technology enables cities to better assess the varying needs of citizens and design suitable solutions. It also has the potential to drive meaningful public participation. For instance, it can act as the foundation for creating and evaluating project plans and budgets that are open for public debate among various stakeholders which can then help decision-makers comprehensively evaluate the returns on investments.

We recommend that the efforts to make data available, accessible, and disseminated be accompanied by a culture of GIS based data-driven decision making. This in turn will let cities leverage the inbuilt capacity to make longer-term decisions including informing public investments. We showcased one such example in the transport sector by estimating the last mile in Mumbai.

This is true for other urban sectors as well. For example, open spatial data on city master plans can inform public debates and participatory planning efforts (Parasa & Pachisia, 2021). This can in turn help decision-makers prioritise specific aspects of plans and adopt course correction measures.

As our cities evolve with different technologies including the Internet of Things (IoT), Geographical Information Systems, blockchain, collaborative computing, and artificial intelligence, they will eventually move towards building new governance tools, such as digital twin cities. To make that journey possible, it is key that urban local bodies start using city data for such long-term policy decisions.

Limitations and Scope for future research

This paper analysed the urban open data landscape in India, both from a people and data perspective. It analysed the different stakeholders present in the ecosystem, in terms of their power and interest in developing the space. Using case studies, it discusses three major data aspects of the current ecosystem that need improvement, namely the lack of city open datastores, data accessibility, and leveraging GIS for decision-making, and demonstrates how such improvement can enable evidence-based decision-making. While these insights are key to developing the urban open data ecosystem, certain limitations need to be discussed that provide scope for future research.

First, and perhaps most importantly, we do not go into an in-depth analysis of the data privacy safeguards that need to accompany the collection, integration and dissemination of such data. To that end, more research on the drawbacks of releasing urban data publicly should be conducted. Moreover, we do not look in detail at the entire open data ecosystem since it was ambitious for this paper. More work on each aspect of this ecosystem needs to be undertaken to ensure that the economic potential of Indian cities can be unlocked while protecting the personal privacy of its residents.

Next, we believe that we have only scratched the surface in terms of analysing the motives of the various stakeholders within the urban open data community since we rely on a review of literature, media publications, and official websites of the initiatives discussed to identify the stakeholders and their role in the urban open data ecosystem. Primary research is essential to better understand the current challenges and incentives of city-level officials across Indian cities in opening their data. For smart cities, consultations with the CDOs can offer insights into the challenges faced by them. This can inform activities for capacity building and the creation of incentives, thereby improving the implementation of programs such as the Data SmartCities, and impacting the open data ecosystem in cities, at large.

The same holds for individual cities as well since they often face idiosyncratic problems that will not have a standardised solution. Finally, like our stakeholder analysis, we recognise that there are a vast array of data issues and potential use cases of urban open data that need to be explored via further research.

Ultimately, the long-term development of the urban open data ecosystem in India is key to unlocking the potential of its cities, its de-facto engines of economic growth, as more citizens flock to urban areas in the coming decades.

References

1. Banerjee, A., Enevoldsen, Nils, Pande, R., & Walton, M. (2019). *Public Information is an Incentive for Politicians: Experimental Evidence from Delhi Elections.* Yale. https://economics.yale.edu/sites/default/files/delhivoter_shared-4_ada-ns.pdf
2. Bhansali, A., Pachisia , H. V., Kwatra, N., Imad, S., Ganapathy, S., & Sippy, T. (2022, May 26). *Covid-19: The mumbai story*. Artha Global. https://www.artha.global/covid-19-the-mumbai-story
3. Bhattacharjee, Subhadeep, Kanuri, C., Maiti, S., & Chaitanya, S. R. C. (2020, June 23). *Feeding into metro rail: new mobility services for last mile commute.* WRI INDIA. https://wri-india.org/blog/feeding-metro-rail-new-mobility-services-last-mile-commute
4. Centre for Monitoring Indian Economy. (n.d.). Centre for Monitoring Indian Economy. https://www.cmie.com/
5. Chattapadhyay, S. (2014). *Opening government data through mediation: Exploring the roles, practices and strategies of data intermediary organisations in India.* World Wide Web Foundation. http://webfoundation.org/docs/2017/09/India-sumandro_oddc_project_report_0.pdf
6. Covid19India (2021). Covid19India. https://www.covid19india.org/
7. Crime in India (n.d.) Crime in India. https://crime-in-india.github.io/about/
8. D'Addario, J. (2020, January 30). *Seven reasons why businesses should be sharing data.* The Open Data Institute. https://theodi.org/article/seven-reasons-why-businesses-should-be-sharing-data/
9. Data Pitch initiative (2019). Data Pitch initiative. https://datapitch.eu/
10. Decidim Barcelona.(n.d)Decidim Barcelona.https://www.decidim.barcelona
11. Deloitte. (2017). *Assessing the value of TfL's opendata and digital partnerships.* https://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf
12. Deng, T., Zhang, K., & Shen, Z.-J. (Max). (2021). *A systematic review of a digital twin city: A new pattern of urban governance toward smart cities.* Journal of Management Science and Engineering, 6(2), 125–134. https://doi.org/10.1016/j.jmse.2021.03.003
13. Esri India. (n.d) *Gis mapping software, spatial data analytics & location platform.* https://www.esri.in?aduc=social&adum=internal&utm_Source=social
14. ETGovernment. (2020, January 27). *Data is becoming increasingly significant for efficient governance: Hardeep Singh Puri - ET Government.* https://government.economictimes.indiatimes.com/news/digital-india/data-is-becoming-increasingly-significant-for-efficient-governance-hardeep-singh-puri/73666449
15. G, A. (2018, December 17). Iisc launches Indian urban data exchange for smart cities in bangalore. REDACT. https://medium.com/redact/iisc-launches-indian-urban-data-exchange-for-smart-cities-in-bangalore-dc43f89309c1
16. Global human settlement (n.d.)—*Ghsl homepage—European commission.* https://ghsl.jrc.ec.europa.eu/
17. Global Smart Cities Alliance (n.d.) *Open Data – gsca v2.* https://globalsmartcitiesalliance.org/?p=463
18. Gurin, J., Bonina, C., & Verhulst, S. (2019). *Open data stakeholders: Private sector.* Zenodo. https://doi.org/10.5281/ZENODO.2677777
19. IDFC Institute. (2021). *Data from BMC's COVID-19 Response War Room Dashboard.* GitHub. https://github.com/IDFC-Institute/mumbai-covid-data
20. IDFC Institute. (2018). SATARC Survey. https://www.idfcinstitute.org/satarc/main/
21. E-Gov Foundation. (n.d.). E-Gov Foundation. https://egov.org.in/
22. EGovernments Foundation. (n.d.) Innovation ecosystem | ecosystem. https://egov.org.in/innovation-ecosystem/
23. HowIndiaLives. (n.d.). HowIndiaLives. https://howindialives.com/
24. IndiaStat. (n.d.). IndiaStat. https://www.indiastat.com/
25. Jha (2022, February 8). *National Budget 2022-23: What Is in It for Urban Local Bodies?* Observer Research Foundation. www.orfonline.org/expert-speak/national-budget-2022-23-what-is-in-it-for-urban-local-bodies/
26. Johnson, J. (2012). *Cities: Systems of systems ofsystems.* Complexity Theories of Cities Have Come of Age (pp. 153–172). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24544-2_9

27. Justice Hub. (n.d.) Justice Hub. https://justicehub.in/
28. Kant, A. (2022, April 26). *Cities will drive India's growth, but only if we build right* The Times of India. https://timesofindia.indiatimes.com/india/cities-will-drive-growth-but-only-if-we-build-right/articleshow/91104853.cms
29. Kwatra, N., & Parasa, R. (2021, July 19). *COVID-19 bed management in Mumbai.* IDFC Institute. https://www.idfcinstitute.org/blog/2021/july/covid-19-bed-management-in-mumbai/
30. MCGM. (2016). *Comprehensive Mobility Plan.* https://portal.mcgm.gov.in/irj/go/km/docs/documents/MCGM%20Department%20List/Roads%20and%20Traffic/Docs/CMP%20for%20Greater%20Mumbai/Executive%20Summary/Executive%20Summary.pdf
31. MCGM. n.d. *BMC on Maps.* https://portal.mcgm.gov.in/irj/portal/anonymous/BMC-on-Map-NearMe-Amenities-Facilities
32. MMRCL. (n.d.). *Mumbai Metro Line 3 Project Route And Stations.* https://www.mmrcl.com/en/project/project-route
33. Mitra, A., & Mehta, B. (2011). *Cities as the engine of growth: Evidence from India* Journal of Urban Planning and Development, 137(2), 171–183. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000056
34. Mobility Data Space (n.d.) Mobility Data Space. https://about.dataspace.mobi/#/
35. MoHUA. (2022, January 17). *SCM, MoHUA launches "Open Data Week."* https://pib.gov.in/pib.gov.in/Pressreleaseshare.aspx?PRID=1790466
36. MoHUA. (n.d.). *Designing a City Data Policy, A Reference Guide.* *https://smartcities.data.gov.in/sites/default/files/SCMCityDataPolicy_Guidance_Nov2020.pdf*
37. Municipal Spatial Data. (2017). Municipal Spatial Data. GitHub. https://github.com/datameet/Municipal_Spatial_Data
38. NITI Aayog. (n.d.) Managing Urbanisation. https://www.niti.gov.in/index.php/verticals/managing-urbanisation
39. NYC Open Data. (2020). *Open Data For All 2020 Report.* https://opendata.cityofnewyork.us/wp-content/uploads/2020/09/2020_OpenDataForAllReport_Full.pdf
40. Open City. (n.d.) Open City. https://opencity.in/
41. Open Data Chapter. (2018, May 22). *Making Cities Open by Default: Lessons from open data pioneers. International Open Data Charter.* https://opendatacharter.net/making-cities-open-by-default-lessons-from-open-data-pioneers/
42. Open Data Initiative. (2016, September 9). *Case study: Open innovation at transport for London.* https://theodi.org/article/case-study-open-innovation-at-transport-for-london/
43. Open Government Action Plan. (n.d) Glasgow City Council. https://www.glasgow.gov.uk/CHttpHandler.ashx?id=56026&p=0
44. OSM. (n.d.) Open Street Map. https://www.openstreetmap.org/#map=5/21.843/82.795
45. Pachisia, H. V., & Ganapathy, S. (2021, November 1). *Civil society should complement, not become a substitute for the state's lack of data capacity.*IDFC Institute. https://www.idfcinstitute.org/blog/2021/november/civil-society-should-complement-not-become-a-substitute-for-the-states-lack-of-data-capacity/
46. Pachisia, H. V., & Gutta, S. (2021, April 29). *Data can play an important role in helping India recover from COVID-19.* World Economic Forum. https://www.weforum.org/agenda/2021/04/india-s-covid-19-crisis-how-data-can-help-cities-recover/
47. Pachisia, H. V., & Parasa, R. (2021, September 8). *Embed open data principles in Master Plans to make planning more inclusive and participatory.* Data Governance Network. https://datagovernance.org/article/embed-open-data-principles-in-master-plans-to-make-planning-more-inclusive
48. Pimpri Chinchwad Municipal Corporation (n.d.). *Pimpri Chinchwad Data Portal.* http://opendata.pcmcindia.gov.in/
49. Praharaj, S., & Bandyopadhyay, S. (2020). *Understanding the open data challenge for building smart cities in india. In S. Hawken, H. Han, & C. Pettit (Eds.), Open Cities | Open Data (pp. 359–382).* Springer Singapore. https://doi.org/10.1007/978-981-13-6605-5_16

50. Press Information Bureau. (2020, November 4). *Hardeep S Puri Launches Nurturing Neighborhoods Challenge, Data Maturity Assessment Framework Cycle 2, and Training Programme for City Data Officers* [Press release]. https://www.pib.gov.in/PressReleasePage.aspx?PRID=1670032
51. Pullock, R., & Rogers, K. (n.d.). *How the UK Government Saved £4 million in 15 minutes with Open Data.* http://opendatahandbook.org/value-stories/en/saving-4-million-pounds-in-15-minutes/
52. Pune Municipal Corporation (2022). *Pune Open Data Portal.* http://opendata.punecorporation.org/
53. Pune Municipal Corporation (n.d) Pune Web GIS Portal. .https://gis.pmc.gov.in/Account/Login
54. PWC. (2020, January). *Preparing for takeoff: How the government sector can leverage drone technology.* https://www.pwc.in/industries/government-and-public-sector/preparing-for-takeoff.html
55. Rial, N. (2013, May 24). *The power of big data inEurope. New Europe.* https://www.neweurope.eu/article/power-big-data-europe/
56. Sachdeva, S., & Patel, S. (2022, June 13). *Local communities should be equal partners in post-pandemic recovery*. World Bank Blogs; blogs.worldbank.org. https://blogs.worldbank.org/sustainablecities/local-communities-should-be-equal-partners-post-pandemic-recovery
57. Sitharaman (2022, February 1). Budget 2022-23. Ministry of Finance. https://www.indiabudget.gov.in/doc/budget_speech.pdf
58. Smart Cities Portal. (n.d) Smart Cities.https://smartcities.data.gov.in/
59. Smith, F., Fawcett, J., & Musker, R. (2017, December 5). *Making agriculture research data more discoverable, reusable & reproducible.* GODAN. https://www.godan.info/blog-posts/making-agriculture-research-data-more-discoverable-reusable-reproducible
60. Snaith, B., Zimeta, M., & Maddison, J. (2021, January 10). *Data enabled cities. Make Better Use of Datain Your City.* https://theodi.org/article/data-enabled-cities/
61. Solapur Municipal Corporation (2019). *City Data Policy of Solapur Municipal Corporation 2019.* http://www.solapurcorporation.gov.in/downloads/data_policy.pdf
62. Spageo. (n.d.). Spageo. https://www.spageo.co.in/
63. Susha, I., Grönlund, K., & van Tulder, R. (2019, January). Data-driven social partnerships: Exploring an emergent trend in search of research challenges and questions. *Government Information Quarterly*, 36(1), 112–128. https://doi.org/10.1016/j.giq.2018.11.002
64. Tata Trusts. (2017, January 20). Tata trusts announces 'the city data for india initiative' in partnership with the world council on city data [Press releases]. https://www.tatatrusts.org/article/inside/tata-trusts-announcesthe-city-data-for-india-initiative
65. Tata Trusts. (2018, May 28). 5 more Indian cities — Ahmedabad, Bhubaneswar, Bhopal, Chennai and Vijayawada – join the data driven governance movement for creating smarter, sustainable, resilient and inclusive cities [Press release]. https://www.tatatrusts.org/article/inside/tata-trusts-and-world-council#:%7E:text=Mumbai%3A%20Five%20more%20Indian%20cities,and%20day%2Dto%2Dday%20operations
66. Tata Trusts. (2019). India's First City Data Officer – Recounting our learnings and exerperiences from Pune. https://smartcities.data.gov.in/sites/default/files/Case1_PuneMC_CDO.pdf
67. The Times of India. (2022, April 29). *Pune Smart City project awarded for open data governance.* https://timesofindia.indiatimes.com/city/pune/pune-smart-city-project-awarded-for-open-data-governance/articleshow/91184409.cms
68. Transport for London. (n.d.). *Our open data. Transport for London.* https://www.tfl.gov.uk/info-for/open-data-users/our-open-data
69. Uber Movement. (n.d.). Uber. https://movement.uber.com/?lang=hi-IN
70. United Nations Department of Economics and Social Affairs. (2018, May 16). *68% of the world populationprojected to live in urban areas by 2050.* https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html
71. United Nations. (n.d) *Cities—United Nations Sustainable Development Action 2015.* https://www.un.org/sustainabledevelopment/cities/

72. Verhulst, S., & Caplan, R. (2015, April). *Open Data: A Twenty first century asset for small and medium sized enterprises.* The Gov Lab https://thegovlab.org/static/files/publications/OpenData-and-SME-Final-Aug2015.pdf
73. vTaiwan. (n.d) Taiwan.https://info.vtaiwan.tw
74. World Bank. (2021, September 30). *New world bank project to support southern indian city of chennai deliver better services to its people.* https://www.worldbank.org/en/news/press-release/2021/09/30/new-world-bank-project-to-support-southern-indian-city-of-chennai-deliver-better-services-to-its-people
75. Zapata, F., & Heeks, R. (2015). *The multiple meanings of open government data: Understanding different stakeholders and their perspectives.* Government Information Quarterly, 32(4), 441–452. https://doi.org/10.1016/j.giq.2015.09.001

Appendices

Appendix A: People | Examples of Stakeholders

1. Public Officials

	Role	Initiative
Ministry of Science and Technology	Sets the nationwide agenda for open data policies	**NDSAP** NDSAP lays down the policy for accessing and sharing non-sensitive data across sectors and ministries in India
MEITY	Sets the nationwide agenda for open data policies and platforms	● **Data.gov.in** It is India's open government data portal which hosts public data from various ministries. ● **Government Open Data License (GODL)** GODL license is open data license for government use that ensures that all users have a permanent right to use the data released by the government under this license.
MoHUA	Formulate nationwide policies particularly related to urban governance and promote data-driven planning in cities	● **Smart Cities Mission** The objective of the smart city initiative is to promote sustainable and inclusive cities by making data-driven decisions. ● **Mobility DataSpace** It was launched in 2020 to serve as a repository for data on transport in cities through the India Smart Cities Fellowship program which is supported by MoHUA. ● **India Urban Observatory**

		It was launched in 2020 and collates data from various sources to generate insights through analytics for cities, academia, industry, and governments. ● **India Urban Data Exchange (IUDX)** IUDX is an initiative housed at the Robert Bosch Centre of Indian Institute of Science (IISc), Bangalore (G.A, 2018). The initiative enables data exchange between various city departments, government agencies, citizens, and the private sector.
NITI Aayog	Promotes data-driven planning in cities	● **Managing Urbanisation** It is a vertical that looks at providing data-based policy for urban planning, development, and management.
National Crime Records Bureau (NCRB)	Official Source of crime data at the national level	● **Crime in India** It showcases the number of offences reported under various sections at the state level and for 50 cities.

2. **Academia**

	Role	**Initiative**
Indian Institute of Science, Bangalore (IISc B)	1. Launch and maintain open data initiatives. 2. Conduct research and develop guidelines for various initiatives.	● **India Urban Data Exchange (IUDX)** (As mentioned above)
Institute of Information Technology, Delhi (IIIT-D)		● **Open static and real-time GTFS data for Delhi** GTFS is a joint initiative by Indraprastha Institute of Information Technology (IIIT-D), the Delhi government and private companies. This initiative has provided access to Delhi's real-time GTFS feeds for buses.

National Institute of Urban Affairs (NIUA)	Advocacy and technical advisory to the MoHUA	● **IUDX (Mentioned above)** ● **Smart Cities Mission (Mentioned above)** ● **India Urban Observatory (Mentioned above)**

3. **Civil Society**

	Role	**Initiative**
Datameet	Analyse data initiatives and create best practices for metadata standards.	● **Opencity.in** It is an urban data portal with 534 Datasets and 1323 Documents. It aims to foster data-driven decisions. ● **Municipal Spatial Data** Collation of spatial Data of Municipalities (Maps) sourced from different government websites.
Oorvani Foundation	Same as above (something around building a platform for publishing city data)	● **Opencity.in (Mentioned above)**
IDFC Institute (now Artha Global)	Generating data through surveys Making data more accessible	● **SATARC** Safety Trends and Reporting of Crime (SATARC) is a survey that maps out the extent and nature of the crime, satisfaction with the police, and perceptions of safety. ● **COVID-19 City Databases** A database created by converting the daily bulletins by BMC into the form of PDFs and images.
Agami, CivicDataLab	Building a platform to aggregate crime data publicly available, thereby increasing its accessibility	● **JusticeHub** Open-source platform for data related to the Indian legal and justice system.

Individuals	Making data more accessible	● **OSM** OpenStreetMap is built by a community of mappers that contribute to and maintain data about roads, trails, cafés, railway stations, and much more, all over the world. ● **Crime in India (Mentioned above)**

Appendix B: Data and Case Studies

1. Data Accessibility and Quality in Pune's Open datastore

The hurdles faced include difficulty in finding the right data platform to access the data from and finding the relevant data within the platform, due to unoptimized search. Further, data accessibility and quality also suffered due to the format in which the data was shared, metadata provided and completeness of the data. Finally, lack of information on methods or processes through which the data was collected also impacts its accessibility and thereby further reuse.

a. Multiplicity of platforms and duplication of efforts resulting in data scattered across platforms

Pune

Spatial data for facilities in Pune is available in two portals: 1. Pune datastore and 2) Pune City Map (Pune Municipal Corporation, n.d.), an interactive GIS portal hosted by PMC. It showcases 31 spatial layers for facilities in the city across 10 categories as of **7th April 2022**. However, these layers are not downloadable directly from the portal for further analysis. It is also unclear if these datasets are the same as those available on the datastore.

Pimpri Chinchwad

PCMC, a neighbouring municipality to Pune, also self-hosts its open data platform or 'datastore'. While the portal had a list of facilities for the Health Infrastructure in the city along with attribute information (such as an address, facility type, number of beds, doctors, and nurses at each facility), it had no location (i.e., latitude and longitude) information. However, the smartcities.data.gov.in portal hosted the location information for the PCMC hospitals, but this dataset didn't contain other attribute information that was available on the datastore for each facility. Also, we couldn't find any spatial data for schools in PCMC on either of these two platforms.

b. Challenges in the Pune Open Datastore Platform

Finding data on Pune open datastore

The process of finding spatial data on facilities on the datastore was cumbersome. In our study, we found that spatial datasets on facilities of the datastore could be discovered on the platform in three ways: 1. On the 'Department' page, selecting 'GIS', 2. On the 'Datasets' page by performing a keyword search for 'Location', 'Latitude'/ 'Longitude', 3. On the 'Datasets' page use a tag filter for tags such as 'amenity', 'bank', 'ATM', 'park' and 'office'. However, while there were some overlaps in the results that the three different searches yielded, none of the three searches individually yielded an exhaustive result that displayed all the spatial datasets on facilities available on the portal. This shows that the users can miss some of the datasets if they do not perform all three searches.

Data and metadata challenges on Pune open datastore

While all the datasets on the Pune datastore are freely downloadable, some of the datasets containing spatial data were not readily amenable to further use and required further processing before they could be imported into a GIS platform for analysis. The latitude and longitude data were not stored in a uniform format across the spatial datasets. The datasets also did not come with a description of the methods used to generate these datasets or the time during which these datasets were created. Most datasets did not feature a sub-categorisation and ownership information of the facilities. For example, in the hospital's datasets, it was unclear if all the hospitals listed were run by PMC or were privately owned and maintained. Similarly, the information on whether some of them functioned as PHCs, CHCs, general hospitals, etc was also missing. These issues also persisted in the data available on the GIS Portal of the city. Such issues limit the user in reusing the dataset as they cannot be fully certain of the contents of the dataset.

2. Last Mile Connectivity and Metro-3 Line in Mumbai

As a first step, we calculate distances between the point locations of the facilities (schools and hospitals) and their nearest train station. Then, we find the average of all such distances aggregated for each ward. Finally, we then measure the change in this distance due to the under-construction metro 3. It's important to note that we calculate Euclidean distances for the purpose of our study. Since these are much lesser than the actual distances people must travel via the roads, they can only be used as a proxy for the actual distances (Combes & Lafourcade 2003, 2005).

The hospitals and schools in Malad, Kandivali and Borivali are farthest (>1.5km) from a train station. These are followed by the eastern suburban wards of Chembur East and West, Mulund and Bhandup and the western suburban ward of Goregaon (>1-1.5km).

We find that the largest drop in the average distance to the nearest train station will be in the southern parts of the city such as Elphinstone (58%) followed by Colaba (48%), Khar (32%), Dadar (27%) and Marine Lines (23%). A decrease of <20% is observed in the suburban wards of Andheri East and West, Goregaon, Bhandup, Kurla and some of the southern wards – Byculla, Grant Road and Sandhurst Road.

About the authors

Harsh Vardhan Pachisia is a former Senior Associate at Artha Global, where he also led the data science unit. His research focuses on topics of urban development, data science and governance. Prior to joining Artha Global, he worked as an Economics Teaching Fellow at the NYU Stern School of Business. Harsh has also been a Research Assistant to several NYU professors analysing topics such as 'Shadow Banking around the World' and 'Social Media Entrepreneurship'.

Anushka Bhansali is a former Associate at Artha Global. Her research focuses on behavioural sciences, public health, and urban development. Prior to joining Artha Global, she interned with the Corporate Economic Cell at Aditya Birla Group and United Nation Development Programme. She graduated with a B.A in Economics degree from Jai Hind College in 2019.

Rajeswari Parasa is a former Associate at Artha Global, where she worked on data science. Her interest lies in building open source tools for enhancing data-driven governance in Indian cities. Prior to joining Artha Global, Rajeswari was a Product Manager at a tech startup, Lakeer Foundation, where she worked on building open-source tools for visualising and interacting with city-data. She was an Urban Fellow at IIHS, Bangalore.