सत्यमेव जयते

NITI Aayog

GLOBAL MOBILITY SUMMIT
SEPTEMBER 7 & 8, 2018

DATA-DRIVEN MOBILITY

IMPROVING PASSENGER TRANSPORTATION THROUGH DATA

Authors and acknowledgements

Authors

NITI Aayog
Pranjal Dubey
Shikha Juyal
Ishita Kandra
Abhishek Saxena
Shweta Sharma
Anil Srivastava

Rocky Mountain Institute
Akshima Ghate
Emily Goldfield
Robert McIntosh
Greg Rucks
Samhita Shiledar
Clay Stranger

Independent contributor
Anand Shah

Contacts

NITI Aayog: adv.niti@gov.in
www.niti.gov.in

RMI: ocs@rmi.org
www.rmi.org

Acknowledgements

The authors would like to thank Dr. Rajiv Kumar, Vice Chairman, NITI Aayog and Mr. Amitabh Kant, CEO, NITI Aayog for their support that made this report possible.

Suggested Citation

NITI Aayog and Rocky Mountain Institute. *Data-driven Mobility: Improving Passenger Transportation Through Data.* 2018.

Outline

1.0

Introduction

+++

Mobility is a key economic driver: moving people and goods is at the core of a well-functioning and prosperous society. Increasing the efficiency and quality of a country's mobility system helps support a stronger economy and a higher standard of living for its citizens.

Data is an essential piece for unlocking maximum value within a transportation system. Data analytics, and data sharing between organizations, has the potential to create more efficient passenger mobility, as well as allow for optimal designing of transit routes and services, infrastructure, and regulations. Data collection is also essential to enabling the use of a number of emerging technologies in the mobility space, such as blockchain and artificial intelligence. Using data to optimize commutes and transit infrastructure will lead to lower levels of congestion, reduced tailpipe emissions, and less time spent in transit, resulting in cities that are cleaner, safer, better designed, and more economically prosperous.

Maximizing the benefits of data analytics for the mobility sector requires the sharing of data between parties. Companies, public transit agencies, and commuters are generating and collecting huge amounts of transport data. However, this data is siloed between organizations and individuals, and recorded with different standards and formats. Absence of transport data ecosystem is an important concern in India as well. Encouraging and enabling data collection and data sharing will allow linking of currently disparate models of transport and unlock tremendous value within the transportation systems in the country.

Data can enable travellers to benefit from seamless multimodal transportation, and make it as convenient as private vehicle use. This paper describes how data can be used to make passenger mobility more efficient, beginning by outlining the relevant mobility data stakeholders and use cases and then analyzing the flow of data in this system. By making possible such a data-enabled future for transportation, mobility assets can be better utilized and integrated, and India can bolster economic growth while building cleaner, more liveable communities.

2.0

The landscape of mobility data

+++

Mobility data encompasses a wide range of mobility-related information. The landscape of mobility data involves multiple stakeholders ranging from data owners and aggregators to data users. The different types of data owners and beneficiaries, and the network of relationships between members of these two key groups, is a core piece of understanding for crafting value propositions to different stakeholders for varying use cases of data.

2.1 Defining mobility data and related terms

Data can be defined as "representation of information, facts, concepts, opinions, or instructions in a manner suitable for communication, interpretation, or processing by humans or by automated means[1]." Mobility data can include a wide range of data related to transport. This includes data of public transit agencies, private mobility solution providers, and individual citizens. In addition to information about transportation assets and trips, mobility data can also include data that affects or goes in tandem with mobility, such as weather data (can affect traffic, etc.), pollution and air quality data, and traffic violations data.

India's proposed Personal Data Protection Act, described in more detail in section 4.2, defines personal data. According to the proposed Act, personal data refers to "data about or relating to a natural person who is directly or indirectly identifiable, having regard to any characteristic, trait, attribute or any other feature of the identity of such natural person, or any combination of such features, or any combination of such features with any other information." It goes on to define the processing of personal data as "an operation or set of operations performed on personal data, and may include operations such as collection, recording, organisation, structuring, storage, adaptation, alteration, retrieval, use, alignment or combination, indexing, disclosure by transmission, dissemination or otherwise making available, restriction, erasure or destruction."

2.2 Primary stakeholder groups

At the highest level, the three categories of stakeholders most critical to the data landscape are **data owners**, **beneficiaries**, and **government**.

Data owners are companies, organizations, and individuals that produce and own datasets. Beneficiaries are groups and individuals that can benefit from using the data owned by the data owners. In addition to both benefiting from and owning data, government can play a key role in enabling the interactions between these two stakeholder groups and protecting their interests. It should be noted that these categories are not rigid: data can flow from owners to beneficiaries, but it can also flow within each of these categories, and some beneficiaries may also qualify as data owners. There are many cases of business to business data sharing; for example, Uber uses data from Google Maps to provide its own service.

Within the category of data ownership, there is one primary distinction: **public** versus **private** data ownership. Public data is shared openly, irrespective of how it is generated. Private data is kept within an organization and not shared with the public. Data owners may have both private and public datasets. Data generated by private service providers is important in India because of a significant presence of privately-owned or operated transport services that meet the passenger mobility needs in cities.

Case study: Digital Matatus Project[2]

Matatus are the privately owned mini-buses in Nairobi, Kenya. Matatus are very popular because they are cheap and convenient, but have had the same problems as a typical informal transit system: difficulty to access timetables, routes, and stops. The Digital Matatus Project is designed to solve these problems using digitization. Digital Matatus Project is a joint project between Columbia University, The Rockefeller Foundation, and the technology sector in Nairobi, and is focusing on developing a mobile routing application and new transit map for the city using cellphone technology. First, students from the University of Nairobi rode all of the bus route using an app that collected data in a General Transit Feed Specification (GTFS) compatible format. Data points included routes, stops, and visual notations (signs and shelters). The data was then processed and released in the form of a paper map and transit apps. Some of the transit apps include Ma3route, Flashcast, sonar, digtitalmatatu, and matatumap. The City of Nairobi has recognised the importance of the digitization and is using this data to create a new trip planning tool for the city. Learning from the success story of Nairobi, several other cities in Africa are planning to map their informal transit sector as well. This case study may prove a relevant example for addressing similar challenges related to the lack of data for India's informal transit system.

A large portion of data owned by private data owners is generated by the movement of individual travellers. If the mobility service is provided through a mobile application, travellers typically agree to allow the company to collect and own the data that the traveller generates. A significant portion of private mobility service providers in India do not have a digital presence. Despite this, these companies are generating significant amounts of data that can be useful. The main challenge in the latter case is finding out ways to collect the data from these service provider's operations.

Each traveller could also be considered a private data owner. Individually, this has much less value than the aggregated data from many travellers, making companies with large aggregated datasets much more relevant data owners than individual travellers in today's data landscape; however, new technologies have the potential to change this relationship. Blockchain, for example, disintermediates the aggregators of data and could allow travellers to have more control over and more direct compensation for sharing the data they generate.

Data beneficiaries can be categorized into three main stakeholder groups: **travellers**, **cities and governments**, and **researchers**. Each of these beneficiary groups correlates with a set of use cases for mobility data, which are explored in more detail in section 3.

» **Travellers** are any individuals needing to move from their current location to an end destination. In this case, the primary use case for open mobility data would be to optimize the route and mode of transport for increased efficiency and ease of use.

» **Cities and governments** include planners, regulators, and operators in charge of system-level design and operations for a city's transportation network, or policy design and operations at a state or national level. The use cases for cities for mobility data include transportation planning (infrastructure and services planning), improving safety, and city planning.

» **Researchers** include any organizations or individuals conducting research in the area of mobility. Greater access to data allows researchers greater insight into the mobility system to better analyze what is happening and draw more data-based conclusions, which could in turn be of use to policy makers and cities.

Travellers represent the individual benefits of data analytics and data sharing, while cities represent system-level benefits. For each of these beneficiaries, the type of data required to achieve the desired use case varies. For example, a traveller needs data that will help him optimize his trip to his destination, which likely only requires real-time data or potentially short-term projections of his transport options when he chooses to depart. In contrast, a city planner who is designing infrastructure for the future would benefit from historic transit data so that she can examine past trends.

3.0

Mobility data use cases

+++

The use cases for mobility data can be examined from the perspective of each beneficiary category. The type of data and analytics required varies between the use case.

3.1 Traveller use cases

There are many potential use cases that apply to improving transportation efficiency for individual travellers. Many of these can be grouped under the general category of “mobility as a service.”

Mobility as a service

Mobility as a Service, or MaaS, refers to the technology-enabled, on-demand availability of multimodal trip options, including multimodal trip planning and seamless payment. Mobility as a service is an alternative to private vehicle ownership; travellers should be able to order a ride to wherever they need to go, at the time they need and in whatever size or type of vehicle meets their needs.

Under the MaaS paradigm, the entire transportation system functions as a cooperative, interconnected system to meet travellers’ needs through a variety of transport modes. To achieve this, infrastructure, technology platforms, payment, transportation services, and data analysis must be capable of working together[3]. Mobility data—as well as organizations’ willingness to share data—is the key to unlocking MaaS. This paradigm has the potential to benefit both travellers as well as transit providers who share data: travellers can enjoy increased options and trip efficiency; while for transit providers, compiled data can help open up markets and customer bases that are currently disaggregated due to the lack of connection between service providers.

There are several data-supported elements that go into MaaS, several of which are described in more detail below.

» **Multimodal trip planning**
A primary element of MaaS is the ability to see all available modes of transport, and both choose the mode that is most optimal for the situation and be able to easily link various modes of transport to get to the destination. For example, a traveller could go onto a single platform and enter his destination and be shown the best option for getting there out of all available modes, which may include a portion of the trip using one mode and another portion using a different mode. Currently, travellers lack easy access to information about all their transportation options. Mobility data—as well as transit providers' willingness to share data—is essential to enabling multimodal trip planning.

» **Seamless Payment**
Enabling travellers to seamlessly pay the various transportation providers through a single portal will increase the accessibility of transport options and promote multimodal trips. Implementing seamless payment requires collection and integration of transit data, and also relies on transit companies being willing to share data.

India Stack's Unified Payment Interface[4], which allows all bank account holders in India to send and receive money instantly from their smartphones, may uniquely position India to be able to quickly adopt a multimodal transit application with integrated payment.

» **Real-time mode connectivity and optimization**
Mobility data can enable the real-time optimization of travel plans around changing factors, such as weather and traffic, as well as travellers' preferences (e.g. least expensive, shortest time, etc.). For example, an algorithm can take into account traffic data to reroute a traveller around an accident or congested area, or update expected travel time based on weather conditions.

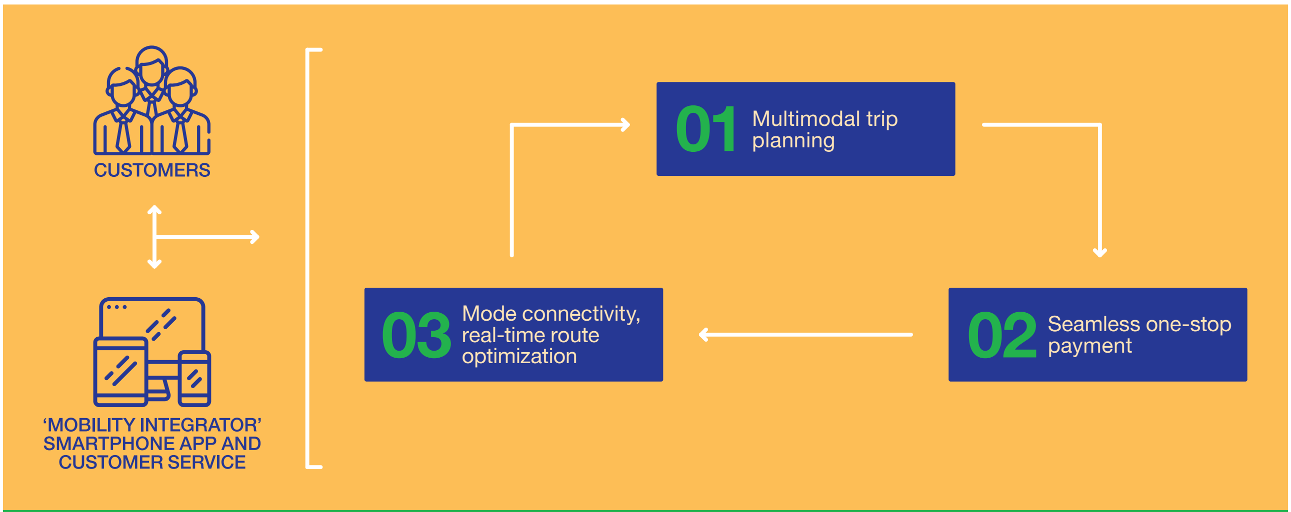


Figure 1: Three interdependent components of "mobility as a service.[5]"

Ancillary trip information

In addition to MaaS, there are a number of other services that access to data can provide to increase the efficiency and ease of a traveller's journey. This could include accessing data on real-time conditions (i.e. traffic, weather, accidents, etc.), information about interesting landmarks along the route, or any other sort of ancillary information to enhance a trip.

Case study: Bangalore's bus tracking application[6]

Figure 2: Screenshot from BMTC's trip planner app.

Bengaluru Metropolitan Transport Corporation (BMTC) has piloted a mobile application using Intelligent Transportation System (ITS) to give fixed and real-time information about BMTC buses to the passengers. Users can easily access the trip planner tool to see the bus timetables and route maps. BMTC buses are equipped with GPS and give a real-time location of buses and estimated time of arrival. The mobile application also gives additional information about the approaching buses at the bus stop, such as bus number and the platform details on which the bus is going to arrive. The app was developed in order to make BMTC bus experience more comfortable for the passengers and promote public transportation as the preferred mode of transport. Similar mobile applications are also being developed in other cities in India, such as Ahmedabad.

3.2 City and government use cases

There are several use cases that apply to optimizing passenger mobility from a systems-level. Cities and governments around the world are realizing the value of using mobility data to improve system safety and optimize transit planning and city design around the efficient movement of people and goods. Some of these use cases are discussed below.

Safety and security

Data can enable improved safety and security within the transportation system in a number of manners. For example, increased access to data allows cities to see where accident hotspots are, thus enabling them to respond more quickly and also understand the issues in those areas. With increased understanding of when and how accidents occur, cities can ensure a greater level of safety for their citizens by responding faster when incidents occur and developing solutions to systemic concerns.

One particular area of concern in India is women's safety in the transportation sector. Many women feel unsafe traveling alone, and frequently avoid using public transport[7]. Improved tracking of vehicles and verification of drivers and vehicles that are deemed safe are some examples of how data can allow women to feel safer using transportation. Safety equipment such as panic buttons, GPS, and CCTVs could be installed in vehicles to improve safety—and tracking of crimes and harassment, should they occur—for women.

Case study: Open Transport Partnership

In 2016, Grab, Easy Taxi, and Le.Taxi formed a partnership with the World Bank and other organizations to share the traffic data from their drivers' GPS streams through an open data license. The three ridesharing companies combined cover more than 30 countries and serve millions of people[8]. The partnership's goal in sharing their data is to provide cities, particularly in developing countries, with the tools to address traffic congestion and plan for better mobility by developing better, evidence-based solutions to road safety and traffic challenges. This unprecedented multi-provider agreement to share data is an important counterexample to the argument that sharing data undermines competitive advantage. A core reason that this initiative was able to launch successfully is that the private companies involved knew exactly what would happen with the data, what type of interface people would see when the data was published, and had assurance that the data would not be used for other purposes.

Transportation, route, and infrastructure planning

Transportation planners can leverage data analytics to better design and maintain routes, public transit, and mobility infrastructure. There are many specific use cases within this category: Analyzing traffic and commute patterns allows planners to understand where to build infrastructure (including non-motorized transport infrastructure) and add transit routes to ease stress in the most highly-trafficked areas. Data analytics can aid planners in minimizing congestion in cities by identifying the primary cause (i.e. poorly-timed signals, insufficient parking, etc.). Data can also aid in monitoring the structural health of transportation infrastructure, such as bridges and overpasses, and related data, such as water and flood data, can inform how and where to build transportation infrastructure.

Real-time system management

Data can aid in the real-time functioning of the mobility system. Operators can remotely monitor the transportation system and manage system operations. Increased access to data will give operators more real-time information that will help them ensure the smooth functioning of the transportation system. One way of collecting this data is by installing sensors to detect traffic and vehicle movement. Additionally, as more and more private companies are collecting data on travellers' movement through GPS, cites are beginning to see an opportunity to partner with these companies to acquire data.

Case study: Waze Connected Citizens Program

Waze, a Google-owned traffic and navigation app, launched its Connected Citizens Program in late 2014 as a two-way exchange of information with cities, with Waze providing city partners with user driving information in return for real-time and advanced notice of construction and road closures. This partnership benefits both parties: Waze sees the program as a way it can grow and improve its services, while city and state governments can easily expand their view of their roads and streets without having to invest in more road sensors and traffic cameras. As part of the Connected Citizens Program, Waze has partnered with over 100 cities around the world, who can now use Waze's data to help with city planning and transportation regulation.

In Rio de Janeiro, for example, the Waze API is completely embedded in the city's Control Center, to help with day-to-day monitoring of road conditions[9]. In Boston, the city's central Traffic Management Center uses Waze's real-time data to change the traffic signals in 550 of the city's intersections as needed to reduce congestion. Washington, D.C. has used data supplied by Waze to aid in the city's "war on potholes[10]." In all of these cases, the cities were able to increase access to data without investing in additional monitoring equipment.

This example also demonstrates that private companies are more willing to share data when the use case is very clear (i.e. city planning and regulating) and the use of the data is transparent.

Enforcement and regulation

Increased access to data affords regulators better visibility into the transportation system, allowing them to improve the enforcement of regulations, and develop new and modify existing regulations to ensure a smoothly functioning system. For example, data can allow regulators to see where infractions incur most frequently and consequently improve their enforcement ability in those areas; it can also aid in the collection of road tolls and parking fees.

3.3 Researcher use case

The researcher use case includes anyone conducting mobility-related research, such as academic institutions and think tanks. Increased access to data gives these groups and individuals a greater ability to conduct in-depth quantitative analysis on the transportation system, in order to draw conclusions about and make recommendations for the mobility system. This allows researchers to provide cities and other beneficiaries with more detailed insights, such as the cause and effect relationships between investments and impact, in order to inform what a city should invest in to improve its transportation system.

3.4 Using data to enable emerging technologies

A number of emerging technologies have potential applications in the mobility sector. Increasing the collection and use of mobility data will help unlock the potential of these technologies. This section examines some of the most notable emerging technologies—big data analytics, blockchain, and artificial intelligence—with respect to their potential applications in the mobility space.

Big data analytics

Data analytics refers to the practice of examining large amounts of data to discover patterns, correlations, and other insights[11]. Data analytics and technologies can allow businesses, governments, and other organizations to analyze datasets and draw conclusions to help them make informed decisions. Big data analytics includes a wide range of specific tools and techniques that can be used to gather insights from data; some specific techniques include data mining, predictive analytics, and text mining.

For mobility, big data analytics can help with planning and managing transportation networks and designing and optimizing services to meet transportation needs. Both the private and public sector can benefit from big data analytics for mobility. Some areas where big data analytics can significantly impact the transportation sector include[12]:

» Optimization of transit schedules by analyzing demand and predicting the impact of maintenance, road work, congestion, and accidents,

» Increased safety and reduced environmental impact,

» Fleet optimization and predictive maintenance through real-time view of fleet operating conditions, statistics around usage and weather patterns, and maintenance cycles, and

» Freight movement and routing optimization.

Blockchain

Blockchain stands ready to revolutionize significant portions of the transportation space and could be particularly applicable in the Indian context. Blockchain allows for more decentralized transactions through a shared, standardized ledger system, which could allow decentralized passenger and freight systems in India to improve their efficiency without the oversight of a central governing body. Some examples of this include:

» **Improved tracking:** Currently most elements of the transportation system—such as freight, vehicles, and jobs—are tracked through a combination of brokers, middlemen, and paperwork. Goods can be lost, vehicles unaccounted for, or jobs duplicated or lost due to lack of accurate tracking. By using blockchain, these items can be tracked on a shared, secured, decentralized, and trusted standardized ledger system[13].

- **Open-source mobility as a service:** Utilization of blockchain supports multiple technologies on a shared, distributed, and transparent platform. Local startups could provide customized services or fleets, and these could be offered through smartphone applications. This would allow companies to provide more unique, customized, and localized services while allowing startups to take advantage of existing platforms and customer bases[14].

- **Secure data logging and tracking:** For vehicle sharing, customers often have a set of preferences and desires they wish in both their vehicle and their fellow carsharing customers. Blockchain enables a customer to have a secure area for these preferences to be logged, as well as previous experiences in carsharing. Through blockchain encryption, this is a permanent, immutable ledger. This could also be used for insurance purposes, as a vehicle could keep a constant log of its experiences in order to document fault in an accident (similar to the "black box" installed in most airplanes for the purpose of facilitating the investigation of aviation accidents and incidents[15]).

Artificial Intelligence (AI)

Artificial intelligence is expected to revolutionize many traditional industries over the next decade. There are a number of areas in which artificial intelligence is impacting the mobility sector, some of which are outlined below.

- **Advanced Driver Assistance Systems (ADAS):** Artificial intelligence can help to automate, adapt, and enhance vehicle systems for safety and better driving, to reduce the potential for human error. These technologies may increase safety and reduce collisions and accidents by alerting the driver to potential problems, implementing safeguards, and taking over control of the vehicle to varying degrees. There are multiple levels of automation in advanced driver assistance systems, the most mild being a system that issues warnings to a human driver, and the most intensive being fully autonomous vehicles.

» **Supply chain management:** In the logistics space, the use of artificial intelligence will lead to more effective distribution of resources in the freight sector as supplies are more effectively moved where they need to go at the right time. Locations and movement of vehicles will be more effectively determined.

» **Infrastructure planning:** As big data becomes more available, cities will start to rely on learning machines and artificial intelligence to process these sets into useful outputs. Artificial intelligence will be used to predict where and how transportation should be built, determine effective mobility corridors, and recommend infrastructure projects.

Pursuit of machine learning in the mobility space

Figure 3: McKinsey 2017 assessment of companies pursuing mobility-related machine learning markets

Companies worldwide are pursuing either new mobility industry areas (such as autonomous driving and predictive maintenance) or old areas with improvement through automation (fleet management, infrastructure planning, supply and manufacturing management, quality assurance, and in-vehicle experience). These companies are traditional mobility companies, startups, and outside industries entering into the mobility space. Given these trends, India could explore machine learning and automation in order to stay competitive on the global scale[16].

4.0

Data acquisition for mobility use cases

+++

This section explores the process for acquiring data, as well as the challenges associated with it, to support mobility-specific use cases. There are several components that play a role in data acquisition, and the following sections explore ways of addressing and minimizing some of the key challenges.

Data acquisition can be done in a primary manner, e.g. through an organization's own sensors or user application, or in a secondary manner, e.g. by acquiring existing datasets from other parties. In the case of primary data collection, organizations often collect data regardless of whether there is an immediate use case for it, under the assumption that owning more data is an asset and may provide useful insights in the future. With secondary data collection, however, it can be helpful to first identify the desired use case before making a request for data.

More and more transit organizations, both public and private, as well as individuals, are collecting significant amounts of transit-related data. The range, scope, and volume of data collection is expanding. However, a significant number of mobility service providers in India still do not collect data, do not have the ability to collect data, or do not operate through digital platforms to facilitate the easy collection of data. Technology can be an enabler in cases of such mobility providers to enable data collection. For example, many cities are using sensors to collect data to improve signal synchronization.

This increase in data presents a massive opportunity to better integrate existing transport systems, optimize transit options to users' needs, and plan and regulate cities to best support mobility patterns. The potential value that mobility data can unlock has led some analysts to refer to data as “the new form of oil” for transport systems. However, there are still many barriers to data acquisition that need to be addressed in order to realize the full potential value of using data analytics to improve mobility.

4.1 Process of data acquisition for specific use cases

When the desired use case for data is known, there are three primary steps for acquiring the necessary data: identifying the sources for the required data; surveying what data is available and identifying what gaps remain; and collecting additional data or acquiring data from other data owners to fill those gaps.

1. Identifying necessary data

The first step is to identify what sort of data is needed to fit the desired use case. This is an essential step; thoroughly analyzing what types of datasets are needed—and conversely, what data is not needed—will save time and allow the organization to make specific asks for data, if additional datasets are required from other data owners.

> ***Example: Route Planning***
> *A city government wants to use data to inform public transit route planning. To do this, a team of transportation planners determines that they want historic data on commute patterns and traffic congestion in order to identify the most highly trafficked routes to determine where new mass transit routes can be developed to ease congestion. They recognize that they do not need real-time data or ancillary data such as weather information, trip fares, etc.*

2. Determining what data is available and what gaps remain

Once the necessary datasets have been identified, the organization can survey which data are already available to them, either through data that they already own or have access to, or through publicly available data. There are already many sources of open data; surveying all of the data that is already available will prevent an organization from collecting redundant data or making unnecessary data requests. Once the organization has determined which of the datasets are already available to them, they can then identify where gaps still exist and what sort of data could be acquired to fill these gaps.

> ***Example: Route Planning***
> *The transportation planners note that they have open access to historic data on the ridership of current public transit options from the public transit agency, and they also have data on traffic patterns from the several traffic sensors that have been installed throughout the city. They then determine that their existing traffic pattern data does not have as much detail as they*

would like, and leaves off a few key areas of the city. They decide to acquire additional data for commute and traffic patterns in the city.

3. Collecting remaining data

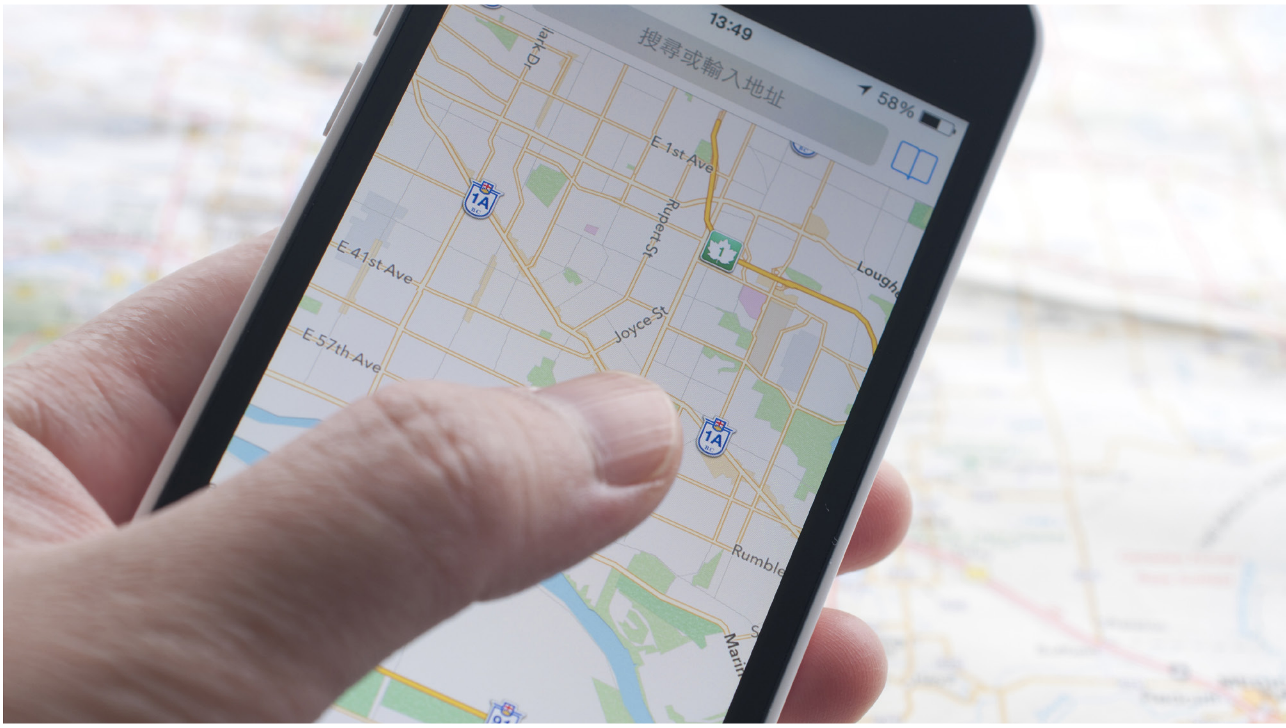
To fill in the gaps identified, an organization or individual has two options:

1. Acquire the data from another data owner: If someone already owns the data needed, one option is to acquire the data from that data owner.

2. Collect the data: If no one already owns the needed data, or if the data owner is unwilling to share the data, then the organization must devise a way to collect the data themselves.

   Example: Route Planning

   The team of transportation planners evaluates the options of installing more traffic sensors to collect their own data on commute and traffic patterns or approach other data owners to acquire data that already exists. They decide to take a combined approach of installing additional sensors, as well as working with shared mobility providers to acquire existing datasets.



4.2 Challenges associated with data acquisition

There are a number of challenges associated with data acquisition,ranging from the quality and availability of data to concerns around privacy and security once data is collected or acquired. The following sections highlight some of these key challenges.

Privacy and data security

As big data becomes more prevalent and necessary in developing effective mobility systems, concerns have been raised about protecting individuals' privacy. Personally identifiable information (PII) is generally considered information that can be used on its own or with other information to identify, contact, or locate an individual person[17]. Individuals are generally concerned about leaks of PII because they can result in negative consequences.

Data does not need to include personal identifiers to be useful in many cases. Metadata—data that describes and gives information about other data and is not attached to a particular individual—can be extremely useful in determining how mobility systems are being used, by whom, and when, to help in determining which systems should be built or supported more effectively. The proposed Personal Data Protection Act of 2018 prohibits the processing of sensitive personal data without explicit consent. Thus, any company or government agency gaining access to data points about mobility users will need to do one—or both—of two things:

» Remove any PII from the data before allowing it to be used publicly.

» If data is to be transferred, ensure the process is completely secure so that PII remains only in the possession of the parties that have been authorized to own or access it.

For the first solution, personal identifiers must be reliably removed from the data so that individuals' privacy is not compromised. Though many datasets include PII, personal identifiers are rarely necessary for transportation planning.

For the second solution, data containing PII (or ideally all data) should be sent through secure channels when it is transferred for use. If PII must be kept and transferred, it should be done in such a way as to eliminate the possibility of access by outside parties.

India's draft data protection and privacy guidelines: Personal Data Protection Act and "A Free and Fair Digital Economy"

In July 2018, the Government of India released the drafts of two major reports on data protection: "A Free and Fair Digital Economy" by the Chairmanship of Justice B. N. Srikrishna, and the Personal Data Protection Act, 2018 by the Ministry of Electronics and Information Technology. The goal of these reports is to introduce standards for data handling, processing, and privacy.

B. N. Srikrishna's report on data protection[18] defines personal data based on its identifiability. The report elaborates: "Identifiability in circumstances where the individual is directly identifiable from the presence of direct identifiers such as names is perhaps uncontroversial and will obviously be included within the scope of any definition of personal data. The definition should also, in addition, apply to contexts where an individual may be indirectly identifiable from data that contains indirect identifiers[19]." The report recommends that a data protection law should be set up, that will be responsible for the enforcement and effective implementation of the definition of personal data and sensitive personal data, legal affairs, policy and standard setting, research, and awareness.

B.N. Srikrishna's report outlines seven key principles for effectively designing a privacy policy. At a high level, the principles are that the policy should be:

1) Technology agnostic
2) Holistic
3) Include language on informed consent
4) Recommend data minimization
5) Assign controller accountability
6) Structure enforcement
7) Include deterrent penalties[20]

The Draft Personal Data Protection Act[21] focuses on the fair and reasonable processing of data. The Draft Act specifies that there must be a clear, specific, and lawful purpose behind data processing and stipulates that only necessary data should be collected. The Act proposes to replace traditional terms such as data controller and data subject (i.e. person whose data is being collected) with data fiduciary and data principal, respectively. The data fiduciary would be responsible for ensuring the personal data quality as well as complete, accurate, and not misleading data processing. Sensitive personal data may be processed on the basis of explicit consent.

The reports are currently under review by the Ministry of Electronics and Information Technology. The government is also seeking comments on the drafts from relevant stakeholders and the public[22].

Poor quality and incomplete data

Data collection in India is insufficient and in some cases the data that is collected is of poor quality or incomplete. A dataset may be considered of poor quality if it fails to provide adequate or accurate information. For example, real-time location data is often inaccurate, collected infrequently, or restricted to only certain services[23]; a transit agency may be inconsistent about stop and route identifiers, or a bus equipped with GPS may have a system that is broken or inaccurate. In order to craft better mobility systems, better data collection will be necessary in India's future.

Acquiring data from private data owners

Acquiring data from private data owners is often one of the biggest challenges in collecting the data needed for a particular use case. These data owners tend to be concerned primarily with jeopardizing their competitive advantage by sharing their private data. The nature of their concern may vary with the use case for the data; for example, data owners often feel less threatened by providing only real-time data to help travellers optimize their transport options, whereas they are more wary of providing historic data to planners and developers. This concern must be addressed by making appropriate and transparent requests for data, as well as making clear the value proposition of sharing data to the data owner.

5.0

Data aggregation

+++

Data aggregation, as a general strategy, is any process in which information is gathered and summarized in certain forms, usually for statistical analysis. Commonly, it is used to obtain more information about particular groups by determining patterns of behavior. It typically involves compiling and anonymizing data from multiple sources, usually for the purposes of observing system-level patterns or understanding broader trends.

In transportation data, aggregation strategies are particularly useful for two reasons: research and privacy purposes. In the modern age of IOT (Internet-of-Things), large volumes of transportation information is easier than ever to collect. The advent of technologies such as GPS, ride-sharing, and navigation and routing programs provides multiple avenues to track individual pathways of movement.

Data aggregation in mobility allows researchers, planners, and private industries to understand larger trends about user profiles of their city. Similar to how a store might market specific products to a certain demographic, data aggregation in mobility enables planners to see how certain user profiles might be using transportation differently.

While granular information is immensely useful for urban planners to understand city behavior, it also raises the question of individual privacy. Data aggregation is a method that enables these city practitioners and developers to access invaluable data that may otherwise not be available, due to the risk of privacy or competitive risk. By aggregating based on certain districts, areas, or types of user profiles, data aggregation anonymizes individuals and removes granularity to such a degree that the user information is now shareable. It is often a compromise between private companies with mobility data or various smartphone companies or routing systems—which contain privacy-sensitive, individual user data—and public city practitioners who are seeking information to better understand their constituents.

Though often a powerful tool, these aggregation strategies can run the risk of failing to provide the information that users need. Private companies are more willing to share their aggregated, rather than granular, data, but the aggregated information may end up obscuring the information that is most necessary for the use case, such as city planning. The goal in data aggregation should be to protect individual privacy while providing useful context to understand and improve the transportation demands of cities.

Case Study: Uber Movement Tool

Uber's Movement tool, launched in January 2017, provides anonymized data to help urban planning in cities around the world[24]. The tool is free and open to anyone interested in using it, and provides travel times between any two points in a city at any time of day. The data provided is an aggregation of many individuals' trips, anonymized and compiled to remove a level of granularity that would violate individuals' privacy or compromise Uber's competitive advantage.

Uber Movement aggregates data in a way that is effective to visualize traffic flows and major congestion points; Uber Movement's data aggregates travel time by city zone, which is useful to understand the impact of major events, from road closures to natural disasters. However, city and transit planners have pointed out that the tool leaves out data that would be most useful to them, such as the most common pickup and dropoff[25] locations in a city. This example represents the conflict between private interests and the public benefits of data sharing. Despite mixed reviews of the tool's usefulness, however, Uber Movement presents a good example of how data aggregation can be used to increase public access to data that otherwise may not be shared at all due to privacy and competition concerns.

6.0

Open data

With a growing population and increasing demand for transport, India stands to benefit tremendously from the collection and sharing of mobility data. Using and promoting open data increases the availability of data to a wider audience.

According to the National Informatics Centre, "A dataset is said to be open if anyone is free to use, reuse, and redistribute it—Open Data shall be machine readable and it should also be easily accessible[26]." Open data is inherently free and available to the public.

Open data helps to enable the many potential use cases for data analytics in the transport sector, as outlined in section 3. The Government of India has taken steps toward increasing the amount of and access to open data in an effort to unlock the many benefits of data analytics. While this supports multiple sectors, the mobility sector in particular has the potential to benefit tremendously from this initiative.

The Government of India, and a number of other countries and cities around the world, host online open data portals with datasets for many sectors, including buildings, education, public safety, and transportation. The goal of these portals is to provide easy access to open data and information in order to "spark innovation, promote public collaboration, increase government transparency, and inform decision making[27]." These portals have been successful in collecting and hosting a wide range of data, and presenting it in an easy-to-use and accessible manner. These central data portals can add a lot of value by aggregating datasets in a single location and making them easily available to any party who can derive value from them. The U.S. cities of Austin and Chicago are two other notable global examples of well-organized, frequently updated, and comprehensive open data portals; both websites have categories for “transportation” with dozens of datasets, ranging from traffic camera and sensor data to non-motorized transit infrastructure maps and plans to public transit ridership numbers.

Case Study: Open Government Data (OGD) Platform in India

The Government of India has launched Open Government Data (OGD) Platform (data.gov.in) to support the Open Data Initiative for nation-wide data sharing. This initiative was launched as part of the National Data Sharing and Accessibility Policy (NDSAP) as well as the Digital India Initiative. OGD platform provides open access to datasets, documents, services, tools, and applications collected by various ministries, departments, and organizations of the Government of India for public use. The main goal of OGD is to provide open access to the data generated through public funds and to enhance transparency, accountability, citizen engagement, collaboration, better governance, decision-making, and innovation[28]. Currently, there are over 2 lakh data points, which have cumulatively been viewed over 17 million times.

7.0 Conclusion

+++

Data has an important role to play in helping India achieve a mobility system that is clean, efficient, and adequately supports the mobility needs of its citizens. To ensure the maximum benefit of mobility data, steps should be taken to collect and share data and ensure that the data collected by different parties is made available as much as possible. Effective communication between data owners and potential beneficiaries is at the core of reaching this outcome; to effectively communicate and collaborate, all stakeholders involved must understand the landscape of the system and the motivations and risks of the various parties involved.

With the rapid pace of technological developments and urbanization, India should be developing and enhancing its transportation system with an eye to the future: as the nature of cities and transport changes, mobility plans need to be designed to be flexible and adaptable to keep up with an evolving landscape. Building a comprehensive practice of data collection and sharing in India will form a strong foundation to support new technologies and innovations, such as blockchain, in the mobility space.

Thoughtfully constructing a supportive framework for collecting and sharing mobility data will enable India to dramatically improve the efficiency and strength of its mobility system as well as urban planning and regulation, resulting in communities that are cleaner, safer, and better support the needs of their citizens.

8.0

Citations

1 Government of India, "Personal Data Protection Bill 2018," 2018. http://meity.gov.in/writereaddata/files/Personal_Data_Protection_Bill%2C2018_0.pdf

2 Digital Matatus, 2018. http://www.digitalmatatus.com/map.html

3 Carlin, Kelly, Bodhi Rader, and Greg Rucks, "Interoperable Transit Data: Enabling a Shift to Mobility as a Service," Rocky Mountain Institute, October 2015. https://rmi.org/wp-content/uploads/2017/03/Mobility-InteroperableTransitData-Report.pdf

4 "About UPI API," IndiaStack. http://indiastack.org/upi/

5 Carlin, Kelly, Bodhi Rader, and Greg Rucks, "Interoperable Transit Data: Enabling a Shift to Mobility as a Service," Rocky Mountain Institute, October 2015. https://rmi.org/wp-content/uploads/2017/03/Mobility-InteroperableTransitData-Report.pdf

6 "BMTC: Easy Travel Information Planner," Google Play, 2018. https://play.google.com/store/apps/details?id=com.bmtc.mybmtc&hl=en

7 Sonal Shah, Kalpana Viswanath, Sonali Vyas and Shreya Gadepalli, "Women and Transport in Indian Cities," ITDP and Safetipin, December 2017. https://smartnet.niua.org/sites/default/files/resources/171215_women-and-transport-in-indian-cities_final.pdf

8 “The World Bank Launches New Open Transport Partnership to Improve Transportation through Open Data,” The World Bank, December 19, 2016. http://www.worldbank.org/en/news/press-release/2016/12/19/the-world-bank-launches-new-open-transport-partnership-to-improve-transportation-through-open-data

9 Stern, Noah, “Waze Partners With Local Governments to Enhance Mobility,” Government Technology, February 12, 2016. http://www.govtech.com/dc/articles/waze-partners-with-local-governments-to-enhance-mobility.html

10 Ibid.

11 “Big Data Analytics: What it is and why it matters,” SAS. https://www.sas.com/en_us/insights/analytics/big-data-analytics.html

12 “How Big Data & Advanced Analytics Will Change Transportation,” Datafloq, December 2016. https://datafloq.com/read/big-data-advanced-analytics-transportation/1747

13 Winnesota, “How Blockchain is Revolutionizing the World of Transportation and Logistics,” 2017. https://www.winnesota.com/blockchain

14 Cohen, Boyd, “How to power shared mobility startups with blockchain technology,“ Shareable, March 2018. https://www.shareable.net/blog/empowering-shared-mobility-startups-with-internet-of-mobility-blockchain-technology

15 Sandner, Phillip, “Analysis of Blockchain Technology in the Mobility Sector,” Medium, April 2018. https://medium.com/@philippsandner/analysis-of-blockchain-technology-in-the-mobility-sector-1078e429615f

16 Cornet, Andreas et al, “The road to artificial intelligence in mobility—smart moves required,” McKinsey & Company, September 2017. https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/the-road-to-artificial-intelligence-in-mobility-smart-moves-required

17 “Personally Identifiable Information (PII),” Cornell Law School. https://www.law.cornell.edu/cfr/text/2/200.79

18 Srikrisna, B.N. “A Free and Fair Digital Economy Protecting Privacy, Empowering Indians,” 2018. https://www.medianama.com/wp-content/uploads/Data_Protection_Committee_Report-comp.pdf

19 Mandhani, Apoorva, “Govt. Releases Justice Srikrishna Committee Report And Draft Personal Data Protection Bill, 2018,” July 2018. https://www.livelaw.in/justice-srikrishna-committee-releases-report-and-personal-data-protection-bill-2018/

20 Singh, Pratap Vikram,"Seven Principles for Data Protection: BN Srikrishna Committee," Governance Now, 27 November 2017. http://www.governancenow.com/news/regular-story/seven-principles-for-data-protection-bn-srikrishna-committee-

21 Government of India, "Personal Data Protection Bill 2018," 2018. http://meity.gov.in/writereaddata/files/Personal_Data_Protection_Bill%2C2018_0.pdf

22 Agarwal, Surabhi, "Government seeks public comments on Data protection bill; September 10 deadline," The Economic Times, August 2018. https://economictimes.indiatimes.com/tech/internet/government-seeks-public-comments-on-data-protection-bill-september-10-deadline/articleshow/65404228.cms

23 Crane, Jackson and Greg Rucks, "A Consortium Approach to Transit Data Interoperability," Rocky Mountain Institute, 2016. https://rmi.org/wp-content/uploads/2017/03/consortium_approach_to_ITD_report2016.pdf

24 Gilbertson, Jordan and Salzberg, Andrew, "Introducing Uber Movement," Uber Newsroom, January 9, 2017. https://www.uber.com/newsroom/introducing-uber-movement-2/

25 Peterson, Andrea and Scola, Nancy, "Data is Uber's business. But protecting it may be its biggest weakness," The Washington Post, November 18, 2014. https://www.washingtonpost.com/news/the-switch/wp/2014/11/18/data-is-ubers-business-but-protecting-it-may-be-its-largest-weakness/?noredirect=on&utm_term=.5b9ef8330e4f

26 "Open Government Data," Government of India Ministry of Electronics and Information Technology. http://sdtckochi.nic.in/OGD.aspx

27 "The official City of Austin open data portal." https://data.austintexas.gov

28 "Open Data," Ministry of Electronics and Information Technology, June 2016. http://meity.gov.in/open-data

सत्यमेव जयते
NITI Aayog

MOVE
GLOBAL MOBILITY SUMMIT
SEPTEMBER 7 & 8, 2018