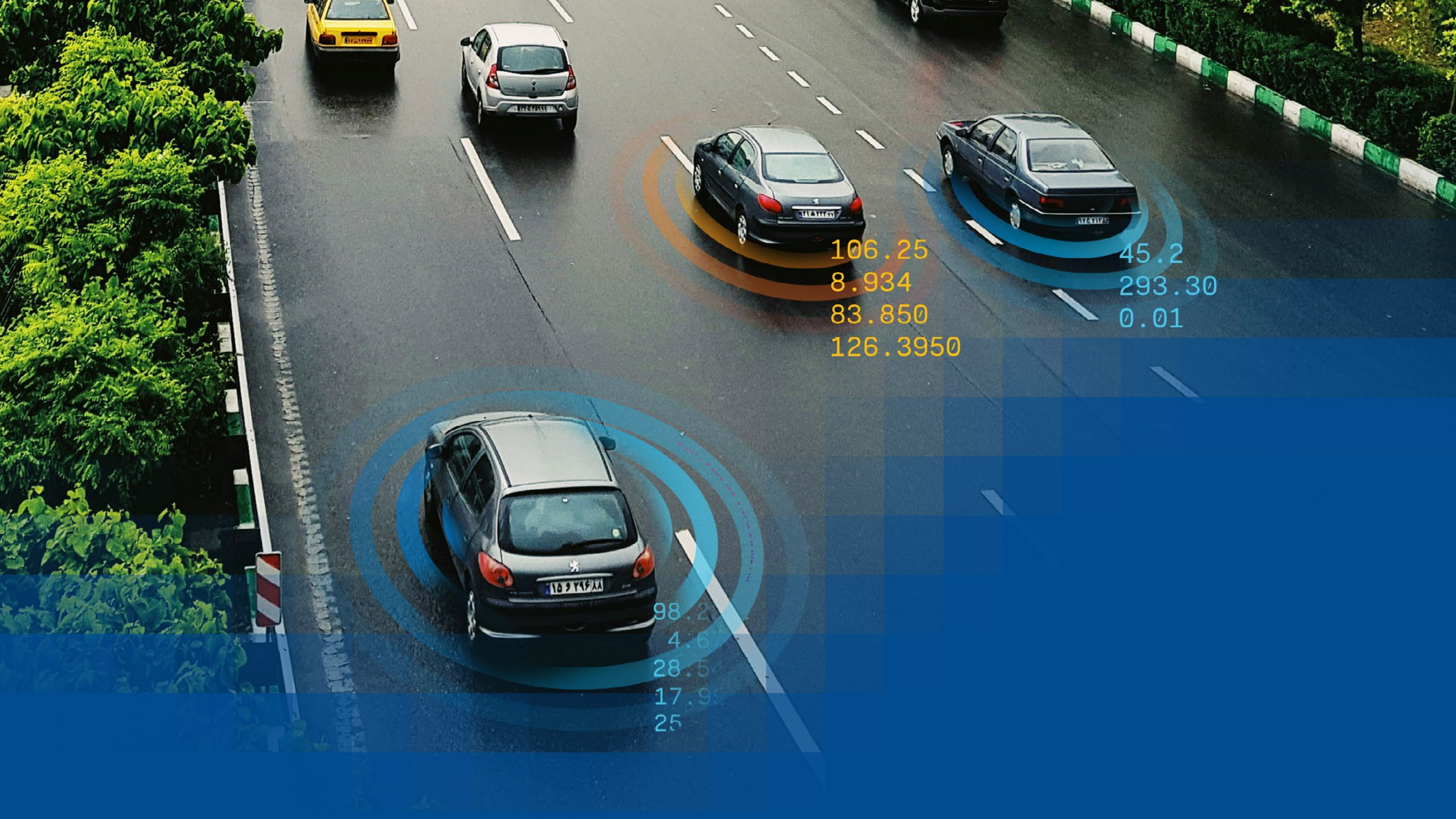


Promoting the Utilization of **Transport Big Data** with **Smart Transport Systems** for **Sustainable Transport** in **Asia and the Pacific**

ESCAP
Economic and Social Commission for Asia and the Pacific

*The shaded areas of the map indicate ESCAP members and associate members.**

The Economic and Social Commission for Asia and the Pacific (ESCAP) is the most inclusive intergovernmental platform in the Asia-Pacific region. The Commission promotes cooperation among its 53 member States and 9 associate members in pursuit of solutions to sustainable development challenges. ESCAP is one of the five regional commissions of the United Nations.

The ESCAP secretariat supports inclusive, resilient and sustainable development in the region by generating action-oriented knowledge, and by providing technical assistance and capacity-building services in support of national development objectives, regional agreements and the implementation of the 2030 Agenda for Sustainable Development.

**The designations employed and the presentation of material on this map do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.*

Promoting the Utilization of **Transport Big Data** with **Smart Transport Systems** for **Sustainable Transport** in **Asia and the Pacific**

Acknowledgements

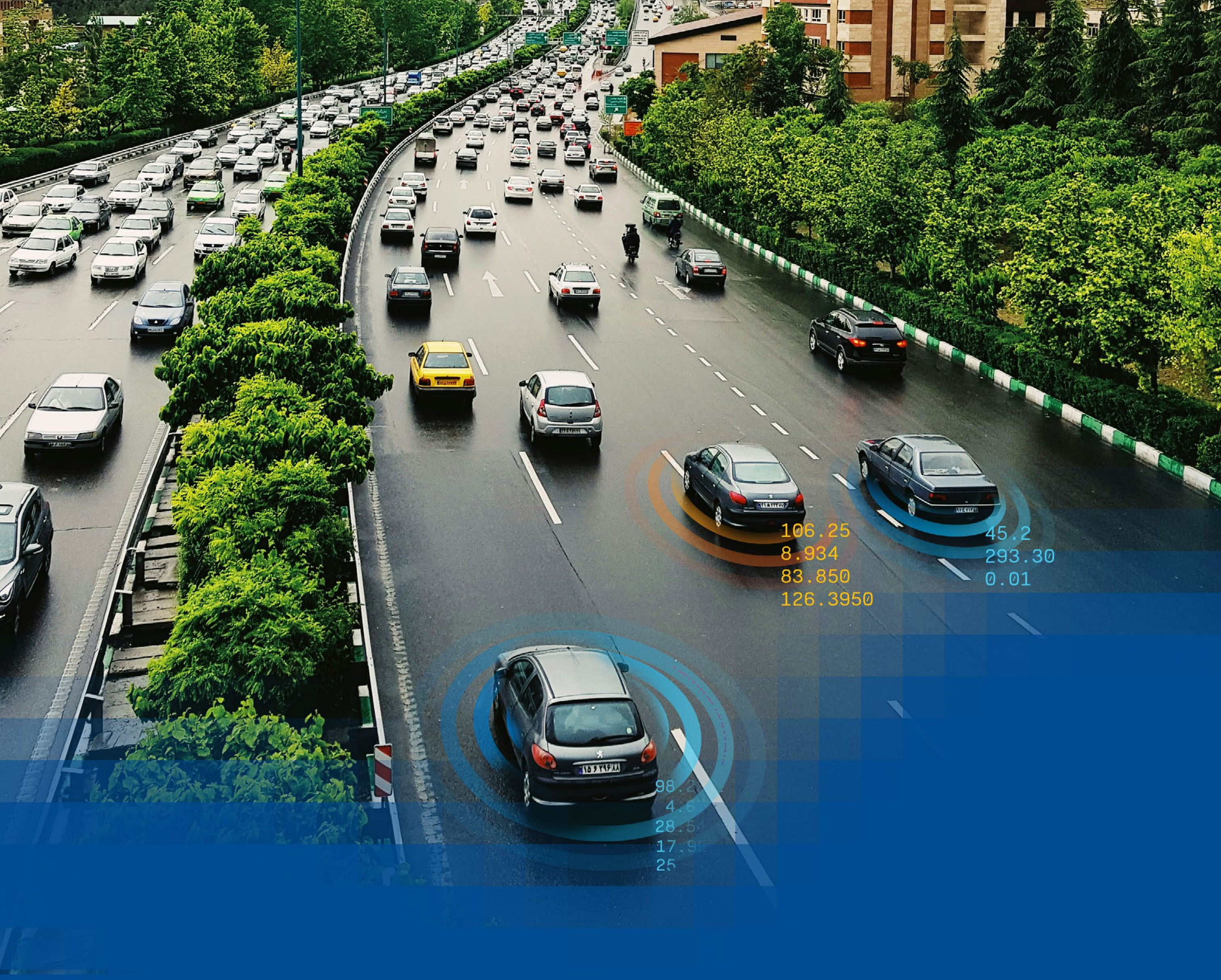
This study was led by Changju Lee, Economic Affairs Officer and Dongik Kim, Expert on Sustainable Transport, and benefited from research support by Hyunmyung Kim, Paul Cranfield, Jiaqi Ma and Mary Sizemore, international consultants, Leong Siew Mun, Erdenechimeg Eldevochir and Intaek Jung, national experts, and Kevin Goggin, economic affairs intern with Sustainable Transport Section. The study was prepared under the overall direction and guidance of Weimin Ren, Director, and benefitted from comments and suggestions from Thanattaporn Rasamit, Section Chief, Transport Division.

The study was conducted as part of the project titled Promoting the Utilization of Transport Big Data from Smart Transport Systems in the Asia-Pacific Region for the Achievement of Sustainable Transport, which is funded by Ministry of Land, Infrastructure and Transport, Republic of Korea.

Contents

Figures

Tables



Chapter 1

Introduction

1.1 Background and purpose

Over the past few years, data has become an important resource in the progress of rapid industrialization. Data plays a more crucial role than traditional resources such as human labor, money, and modes of production, and is contributing to industrial growth. Big Data is an important ingredient for improving Artificial Intelligence (AI) technologies and Information and Communication Technology (ICT) infrastructure and is rapidly transforming the lives of people as well as the national economy. Data, networks, and AI combine to make up the industrial foundation of Data-Network-AI (D.N.A), which has the potential to foster the innovative growth of industries. In other words, a new industry will be established by using data and technology, which will lead to the creation of innovative business models and new services.

Recently, a new economic model with a focus on generating economic value through data has been introduced along with the term, i.e., data economy. The objective of this model is to create new business models and services through data collection, processing, analysis, and utilization, and to motivate the growth of existing industries by increasing their productivity and efficiency. Data economy is closely related to this overall shift, where data is becoming an important resource for the entire industry.

Data economy goes through four procedures to enhance the quality of a data product: 1) data production, 2) data distribution, 3) utilization of businesses, governments, and public institutions, and 4) the use of data-based services by consumers and citizens. Many Big Data-based technologies are being developed by following these procedures, which can then be used to generate valuable information. Along with such developments, more and more people around the world are attempting to analyze and predict industrial and social phenomena. There has been an increasing number of cases where businesses, governments, public institutions, etc., are investigating patterns the relationships between regional issues to use these results in policy decision-making. This is called evidence-based policymaking, and its importance is growing in the public sector.

To apply Big Data to evidence-based policymaking, an in-depth evaluation is needed, such as what types of data are being collected in each country, how they're being collected, and what kinds of analyses and services can be produced from the gathered data. Ultimately, it is important that a smooth pipeline is created, starting with 1) the collection and processing of data, 2) data analysis and modelling, to 3) policy decisions based on analysis results. To accomplish this, the current status and relevant case studies of various transport Big Data and Big Data analysis systems need thoroughly to be investigated and a long-term policy direction fit for the circumstances and characteristics of each country needs to be set.

This report's purpose is to assess the capabilities of governments of countries in Asia and the Pacific and to provide appropriate solutions. There are clear disparities between the levels of development of each country's Big Data industry. This report considers national capability for data economy at an early stage, short- and long-term visions and degree of interference, and the tensions between the public and private sectors as points that should be studied in further depth. This report also aims to further understand the data status of each Asia-pacific country's transport Big Data collection system as well as case studies and analysis systems based on the gathered data. After determining the current situation's limitations and issues through this process, alternatives to better utilize transport Big Data will be offered and short- and long-term goals will be suggested for each country.

Ultimately, this report proceeds by prioritizing the 4 strategies: 1) The collection of evidence-based information and data, 2) the development of a method and structure for the making of evidence-based policymaking guidelines, 3) discussion between experts of countries, and 4) assisting of countries' data-based policies and strategies for the development of smart transport systems.

As such, this report first explains the definition of Big Data and its characteristics that need to be reflected in each country's policies in Chapter 2. Chapter 3 examines each country's transport Big Data database cumulation progress and the status of relevant laws and policies. Chapter 4 describes how Big Data is being collected in each country and draws out implications through case studies with appropriate analyses. Chapters 5 and 6 suggest policy guidelines for each country that enable them to independently create better utilization alternatives through transport Big Data, plans for further improvement, as well as short- and long-term goals.

1.2 Importance of study

Many cities and countries in the Asia-Pacific region are expected to face numerous challenges and obstacles in the process of promoting Big Data-based industrialization and evidence-based policymaking. While some Asia-Pacific countries are already testing and deploying various Big Data technologies mentioned previously, most countries are still in the early stages, and some have not even started. However, Big Data has enormous potential in the transport and mobility sectors. In recent years, innovative ICT technologies such as artificial intelligence, machine learning, and Big Data analysis have been actively utilized in the connected and autonomous driving industry, and as a result, diverse and innovative mobility services have emerged, contributing to the vastness and diversity of transport Big Data.

Hence, with the introduction of smart transport systems, the data revolution has already begun; data has been performing an extremely important role in establishing traffic policies and strategies for cities and sustainable development. Transport data is being used in many aspects, such as evidence-based decision-making, smart transport solutions, policies for regional development, and more.

For example, traffic volume and travel time data can be easily collected through smart transport systems, such as real-time detectors and car navigation systems. Data collected in this way can be used to understand disparities between marginalized groups or areas and users of smart transport solutions. Another example is the use of smart cards when using public transport, which can collect boarding and alighting information of public transport users, as well as personal information such as gender and age. Such high-quality data can be directly utilized in the development of public transport operation strategies and policies, including the revision of bus routes or adjustments of intervals between departures.

However, it is often difficult to consistently collect high-quality transport data in several countries and regions. Data collection infrastructure varies significantly between countries and even within the same country across different regions. These challenges in data collection and infrastructure make it difficult to establish consistent transport policies and plans based on data. Furthermore, even with basic collection infrastructure in place, there are often cases where transport data is discarded due to a lack of awareness and knowledge, and many useful data sets remain unused and unexplored.

The most significant issue is that traffic planners and policymakers lack both understanding and knowledge about how data exists and is being collected in the transport sector, and how this data can be used to solve various problems. As the current stakeholders of the Big Data field are not familiar with or knowledgeable about transport data, there was little utilization of transport Big Data. When considering the immense potential of transport Big Data generated from smart transport systems, the utilization of this data needs to be led by a stakeholder who can contribute to the evidence-based decision-making process. Relating to this idea, with the ultimate goal of sustainable transport in mind, this study intends to become the first step in increasing countries' understanding, knowledge, and capabilities of Big Data.

Chapter 2

Roadblocks in deploying smart transport systems in Asia and the Pacific

2.1 Big Data's definition

Today's world is in the era of Big Data. Many companies and organizations are striving to collect and analyze massive amounts of data to compete in the business world. As a result, there has been active development of open source-based analysis platforms and software frameworks, as well as technologies that collect and analyze diverse data efficiently through hybrid and multi-cloud environments. In addition, with the growing importance of digital transformation, many companies are focusing on leveraging Big Data analysis strategically. This trend is stimulating the growth of the global Big Data market, and it is expected that even more data will be generated in the future.

Due to these factors, the Big Data market has been growing at a fast rate. The size of the world's Big Data market marked 138.886 billion United States dollars (USD) and an average annual growth rate of 10.6 per cent in 2020, and it is predicted to reach 229.423 billion USD in 2025. North America had the highest market share of 30.3 per cent in 2020, and the Asia-Pacific region's growth rate increased to 11.9 per cent with 32.37 billion USD in 2020, forecasted to reach 56.82 billion USD in 2025.

Originally, the concept of Big Data referred to high-volume data analyzed by major companies like Google or the National Aeronautics and Space Administration's (NASA) scientific research projects. However, as Big Data's "Big" is subjective, it is now being defined in numerous ways depending on the subject. The renowned global consulting group, McKinsey, defines Big Data as 'a technology that exceeds the potential of existing database management tools that collect, store, manage, and analyze data, and instead extracts data from a large amount of structured or unstructured datasets and analyzes their results' "Big Data requires an innovative, cost-effective information processing procedure for sharp insight, decision-making, and process automation, and is a high-volume, high-velocity, and high-variety information asset."

While the aforementioned definition from McKinsey focuses on the dataset's volume, Gartner's definition emphasizes the importance of velocity and variety, which is a more universal definition and also the most widespread definition of Big Data known for the concept of 3V. Big Data defined in this manner can be divided into 3 types: structured, unstructured, and semi-structured data:

1 Structured data

Structured data refers to data that can be understood with only numerical values amongst those that follow the standard rules in a database. In most cases, if the data value's meaning is easily understandable and consistent, it can be classified as structured data. Structured data has fixed types of information collected by the database's schema, and the data is organized according to assigned rows and columns. Data that allows the stacking of several samples into a table according to an assigned column, such as various demographic data, test result data, and traffic volume and travel speed data collected through sensors are all examples of structured data.

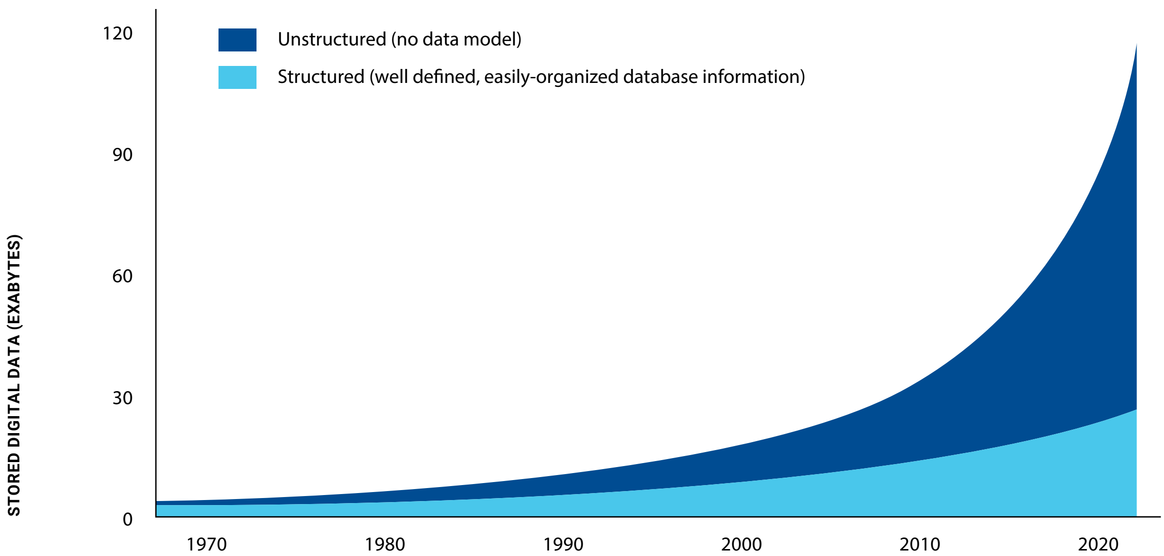
2 Unstructured data

Unstructured data refers to data that does not have a fixed structure or format like structured data. It is free-form in nature and can be created in various formats, making it difficult to understand the value or meaning of the data. Examples of unstructured data include texts, voice files, and videos. The data processing and analysis of unstructured data are challenging due to its diverse nature. Various types of data collected from popular social media platforms and video-sharing websites such as Facebook, Instagram, Twitter, and YouTube are good examples of unstructured data. Structured databases like spreadsheets and traditional databases are not suitable for processing unstructured data. However, with the advancement of technology, it is now possible to analyze unstructured data, enabling the discovery of new insights.

3 Semi-structured data

Data with HTML or XML formats fall under the category of semi-structured data. It isn't a typical database but has a schema. Normally, the database's location for storing data and its schema are separated, forming a mechanism that generates tables and stores data. However, semi-structured data output columns and values onto a single text file. In other words, semi-structured data cannot be perfectly classified from structured and unstructured data per the Mutually Exclusive, Collectively Exhaustive (MECE) principle. Whereas in the primary stages of the data market and industry, structured data was the majority, the recent supply of smartphones and the development of social media have exponentially increased the volume of unstructured data. As unstructured data has very high volume and no structure, technology that aids efficient storage and analysis is crucial. In the past, sensor- or detector-based unstructured data was usually collected in the transport industry as well, but with the advancement of video-based data collection technologies, the types and amount of unstructured data are increasing.

FIGURE 1 Progression of change between the volume of structured and unstructured data



Source: EHL.AI, "Structured and unstructured data", Available at https://ehlaipoint.com/structured-and-unstructured-data/ (Accessed on August 18, 2023).

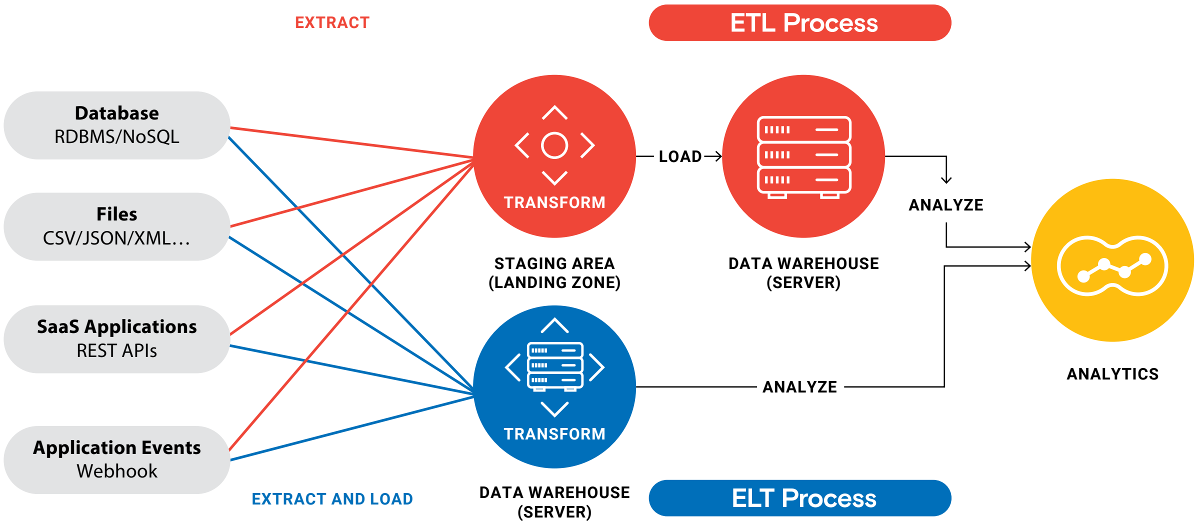
2.2 Components in the Big Data analysis process

The four definitions of components in the Big Data analysis process are as follows:

1 Ingestion (Data Extract and Transform)

Ingestion refers to ETL's Extract and Transformation process, also well known as the Big Data analysis process. Basically, it's a concept that covers the entire procedure of collecting and preparing data. The refinement and transformation of data followed by its extraction are essential processing procedures, and include tasks such as filtering, aligning, aggregating, joining, organizing, and deleting repeated data during data transformation.

FIGURE 2 Big Data Analysis Process Comparison Diagram (ETL and ELT)



2 Storage (data load)

Storage refers to the process of storing the collected data after necessary data processing, which includes extracting and transforming data. Various tasks are involved in data transformation, such as filtering, sorting, aggregation, data join, data cleaning, duplicate removal, and data validation. In the final step of ETL, data is loaded into a data warehouse or data lake. However, in the existing ETL approach, there is a problem that when a large amount of data is generated, a lot of time is spent on the transformation step. To address this issue, ELT (storing data first and then deciding where and how to use this later) is gaining attention. Furthermore, the price of resources such as CPUs, memory, SSDs, and HDDs is becoming cheaper, as well as cloud usage fees, making ELT more prominent. In the storage stage of ELT, data schema and type, as well as data, are directly loaded into the target system.

3 Analysis (data analysis)

In this stage of the Big Data process, data is analyzed, and useful insight is provided to the organization. In Big Data analysis, there are four types: prescription, prediction, explanation, and diagnosis, and AI and machine learning algorithms are utilized.

4 Consumption (data display)

The final component of the Big Data analysis process is Consumption, which involves sharing data with others. Data visualization and storytelling are important tools for effectively sharing insights with non-experts and stakeholders. Data visualization is useful for exchanging ideas and can facilitate decision-making. Therefore, data visualization plays a crucial role in promoting strategic decision-making. Below are some ways in which data visualization can assist with decision-making.

- **Understanding the overall situation:** Since trade, interaction, process, and behavioral data are stored in a system, there is also evidence within the system that demonstrates its performance. In other words, through data visualization, it is possible to recognize the broader context within the data and to understand hidden trends or patterns.

- **Discerning important factors:** Discerning important information involves the ability to identify which data is important and interpreting this data accordingly. To make better decisions based on insights discovered through data storytelling, it is essential to understand the core insights embedded in the data. Through identifying important information, resource allocation optimization can occur, and by understanding which areas to focus on, the effectiveness of the data analysis can be maximized. Therefore, discerning important information plays a crucial role in data storytelling.
- **Information-based decision-making:** By using accurate statistics and visible insight, one can confidently make decisions based on data.

Data storytelling is a technique of presenting data through a narrative that has context. It is similar to traditional storytelling but provides deeper insights and substantiates them through data visualization.

When creating a data story, it is essential to combine three key elements: narrative, visualization, and data interpretation to provide a balanced analysis and an understandable view of the data to maximize its value. The narrative contextualizes the data and analysis results, emphasizing key points. Visualization makes the data easy to understand visually, displaying important patterns and insights. Finally, the data serves as the foundation for analysis, supporting the presented narrative and visualization. This balanced analysis and data interpretation help organizations seeking insights from data to make data-driven decisions. Additionally, such data stories can be easily communicated not only to data professionals but also to the general public, increasing the value of data.

Until now, the definition of Big Data and the definition of each stage in the analysis process was explained. In the next section, the detailed characteristics of Big Data are reviewed to see what benefits Big Data can bring to governments and businesses.

2.3 Big Data's characteristics

As value production became more important, Big Data's 3 primary characteristics, also known as 3V (Volume, Variety, Velocity) turned into 5V, including Veracity and Value. More recently, these characteristics have been further divided into 8V, which adds Validity, Volatility, and Visualization**.**

1 **Volume:** The size and volume of Big Data managed and analyzed by a business.

2 **Velocity:** The speed a business receives, stores, and manages data (ex: the number of certain social media posts or search queries received during a given period).

3 **Variety:** A wide spectrum of data that includes unstructured data, semi-structured data, and basic data.

4 **Veracity:** The "truth" or accuracy of data and information assets that oftentimes influence the credibility of the owners.

5 **Value:** The most important "V" – Big Data's value – from a business perspective. It is usually the recognition of patterns or discovery of insights that leads to effective management, stronger customer relationships, and clearer, more quantifiable business benefits.

6 **Validity:** The effectiveness and relevance of the data that will be used for the intended purpose of analysis.

7 **Volatility:** As Big Data is constantly changing, the data collected from a source yesterday can be different from the data collected today. This characteristic impacts data homogenization.

8 **Visualization:** Visualization refers to the demonstration of insights generated from Big Data through the visual expression of charts and graphs. It is a characteristic that has been spreading recently as Big Data experts began to regularly share their insights with non-experts.

Such characteristics of Big Data are advantageous in that they can be utilized by both private and public organizations in important ways. A few core benefits are explained below.

1 **Strengthening of decision-making:** Big Data supports businesses and organizations so that a decision based on better evidence can be made in a shorter amount of time. Strategies can be adjusted in further detail using search engines, social media platforms, etc. It also aids businesses by identifying unprecedented trends and patterns to prevent errors.

2 **Data-oriented customer service:** The most significant influence that Big Data can have on all industries is in the customer service department. Businesses and governments are increasingly replacing their user-feedback systems with data-oriented solutions. These solutions analyze customer feedback more efficiently and assist organizations so that they can provide customer services to consumers accordingly.

3 **Efficiency optimization**: Organizations use Big Data to identify internal issues. Management can be enhanced dramatically by resolving issues with insights generated from Big Data. Big Data has been especially helpful in increasing the efficiency of the manufacturing sector through IoT and robotics engineering.

4 **Real-time decision-making**: By enabling real-time detection of tasks such as stock management at banks and financial institutions, supply network optimization, money laundering prevention, fraud detection, etc., Big Data has transformed numerous fields over time. If this is applied to the public realm, it is possible to immediately make real-time decisions using Big Data-based analysis for the convenience of citizens.

However, an overly trusting attitude towards Big Data utilization should be avoided, because cases where Big Data only consists of useful data is relatively rare. For Big Data to not merely be a massive aggregate of data but actually good data, there could be accurate and efficient data collection and processing procedures, accompanied by a huge effort to analyze collected data, convert analysis results to services, and ultimately connect this process to policy decision-making.

In particular, because each country and region have varying amounts of data as well as different laws regarding open access to Big Data, it is necessary to strengthen the data utilization capacities of each country according to their best available data and policies. Preparing detailed utilization measures per country regarding their data and developing a technology that additionally collects lacking data would be the most feasible and sustainable utilization method of Big Data.

Chapter 3

Current status of transport Big Data

3.1 Evaluation of transport Big Data

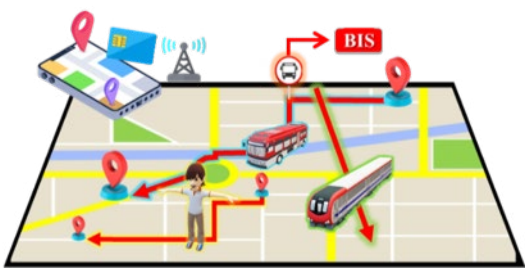
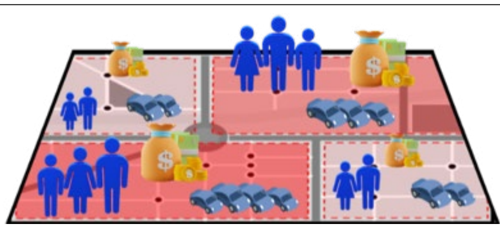
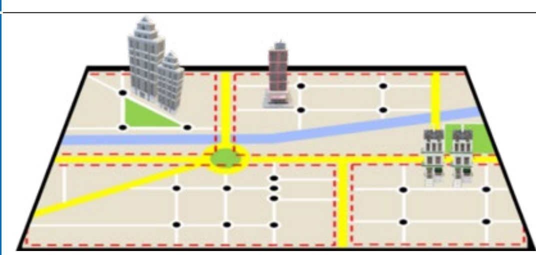
Due to advances in analytical technology, Big Data now plays a significant role in providing valuable information to various fields, such as science, economy, and technology. While it used to be primarily utilized in the private sector, the collection, production, storage, access, and analysis of data are now possible in the public sector due to innovations in information and communication technology (ICT), which has led to an increase in the value of Big Data.

Big Data in the public sector has a significant impact on digital transformation. Since the 2010s, some Asia-Pacific countries have been utilizing Big Data to monitor social and economic indicators to improve public services. Such sector-specific cases are promoting the enhancement of Big Data analysis techniques, policies, and service development in these countries.

The transport sector is an area where Big Data can be effectively utilized. This is because of its large user base and the high utility of real-time information collection and provision. Consequently, various policy and service development demands for Big Data are rising in the transport sector. Transport data can be classified into items collected from human movement data, vehicle movement data, public transport data, and spatial information data. Furthermore, city characteristics data and socioeconomic data are also important for integration and connection with the transport sector.

As the transport sector is primarily under the control of the public sector, Big Data analysis and application can help streamline a country's administrative tasks and contribute to the development and implementation of more sophisticated policies. Therefore, in this section, the most commonly available Big Data in the transport sector are classified into four categories based on their characteristics: 1) transport and mobility data (movement and trajectory), 2) transport and mobility data (operation and facility), 3) socioeconomic data, and 4) basic city and spatial information data.

TABLE 1 List of data by major categories in the transport sector

TRANSPORT AND MOBILITY DATA	**Movement and trajectory**  	**1. People's movement data: Travel Trajectory Data (Individuals or Vehicles)** – GPS Tracking Data, etc. – Web portal Map based optimized Route Information **2. Vehicle trajectory data -Navigation Data** – DTG (Digital Tachograph) Data – DSRC (Dedicated Short Range Communication), TCS (Toll Collection System) Data – CCTV – Drone Video Analytics **3. Public transport data** – Smart Card data – Operation Routes by modes BIS, BMS – OBE (On-Board Equipment), etc. **4. Other transport modes' trajectory data:** – PM (personal mobility) data – DRT (demand responsive transport) data – Bike Sharing data
	Operation and facility 	**1. Traffic Information:** – Traffic Volume – Velocity – Density, etc. **2. Public Transport:** – Bus Network (Stops, Routes) – Bus Management System (BMS) – Railway Network (Stations, Rail routes) – Data about used modes of transport, etc. **3. Public Infrastructure:** – Location of transport facilities – Parking Space location – Traffic Signs, etc.
SOCIO-ECONOMIC DATA**		– Socioeconomic indices (Population, income, number of registered vehicles, etc.) – Conversion coefficient of annual average daily traffic per region (AADT) – National Household Travel Survey (vehicles, passengers, household types), etc.
GLOBAL FUNDAMENTAL GEOSPATIAL DATA***		**4. Administrative Boundaries** – Land Use data – Environmental data – Hydrography data – Elevation data – Multi-temporal data – Traffic network GIS data (TAZ, Link, Node, etc.) – Building location data – PoI (Point of Interest), etc.

* Collected and recorded dynamic travel data including the travel of vehicles, people, and public transport users.
** An index that encompasses the quality of life, economy, society, and environment as a representation of a society's degree of development.
*** Location information about an artificial object that can determine distribution and geographical location displayed on maps.

1 Transport and mobility data (movement and trajectory)

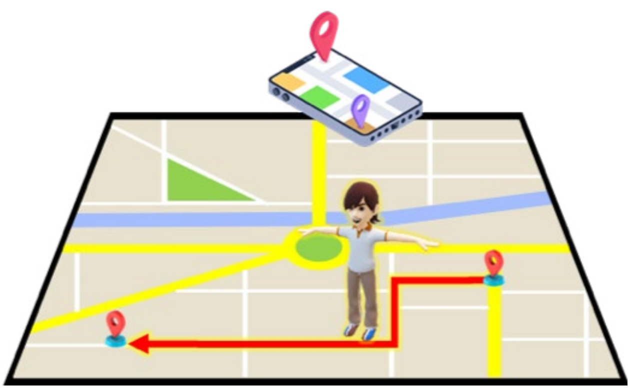
The main subjects of the transport sector's mobility analysis are people, vehicles, and public transport. Transport and Mobility Data (movement and trajectory) typically divides mobility data into 4 categories: (1) people's movement, (2) vehicle trajectory, (3) public transport mode's movement, and (4) other mobility modes' movement.

People's movement data

The trajectory of movement is one of the dynamic travel information collected along with information on pedestrian characteristics, volume of traffic at different times and locations, population flow, and travel times by region. Nowadays, as the amount of people's movement and modes of transport have increased, travel-related information has also increased. In the private sector, location information is typically collected from global positioning system (GPS) in mobile devices used by individual travelers. This data includes transmission and reception information between a smartphone and a base station of a telecommunication company, enabling analysis of the movement characteristics and trajectories of individual travelers. Since cellular data records all travel information of smartphone users with activated devices, and considering the number of subscribers of telecommunication companies, data collection is possible for almost all citizens, resulting in a very high sample rate.

However, directly identifying the characteristics of original data is very difficult, because an individual's personal information needs to be protected. This limits the amount of people's movement information that can be released by the public sector. Typically, only data that is aggregated by origin-destination, administrative units, or within a grid system is made available to the public to ensure anonymity.

TABLE 2 People's movement data



CATEGORY
Mobile Communication Data
Foot traffic Data

Vehicle trajectory data

Individual vehicle trajectories, as well as aggregated travel information, such as traffic volume, average speed, travel ratio, and travel time, are collected to assess the level of crowdedness and calculate road congestion costs. Vehicle movement data is collected using various sources such as Dedicated Short Range Communication (DSRC), Digital Tachograph (DTG), and navigation systems, depending on the vehicle's characteristics. The collection methods also vary depending on the road management authority. In the Republic of Korea, for example, the Korea Expressway Corporation collects individual vehicle trajectories and speed information of vehicles on highways using DSRC and Roadside Equipment (RSE) sensors installed along the roads. The Korea Transportation Safety Authority installs DTG terminals on commercial vehicles such as buses, taxis, and trucks to collect and manage the vehicle's location, speed, direction, and risky driving behavior information on a second-by-second basis.

Private companies that offer navigation services can collect vehicle movement information through navigation or smartphone app-based communication mobile data used by drivers. Therefore, the companies can collect individual travel histories not only on highways but also on general urban roads. Vehicle GPS data collects vehicle location information at a second-by-second basis by using navigation and DTG data and can produce various information such as vehicle travel paths, traffic volume by road sections, queue length by intersection, and travel time.

TABLE 3 Vehicle trajectory data

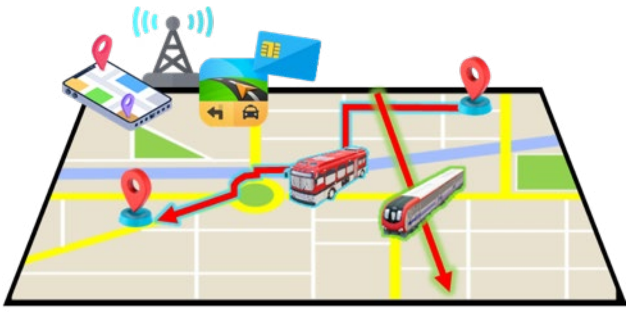
CATEGORY
Mobile Communication Data
DSRC Data
CCTV Data
Drone Video Analytic Data
Navigation & GPS Data
Digital Tachograph (DTG)

Public transport data

Public transport, such as buses and subways, is an essential service for citizens who do not own cars. Public transport services should be a basic service accessible to all citizens and operate as the primary means of transport in cities. The movement information of public transport modes is collected through various methods, the most common data being Smart Card data, also known as public transit card data. Additionally, data on vehicle locations, route operations, passenger usage, and real-time operation history are provided by the operation agencies of each transport mode. Smart Card data consists of information scanned by public transport users, such as the passenger's identification (ID), boarding and alighting stops, boarding vehicle information, and more. This data can be used to produce various information, such as crowding information within public transport vehicles, boarding and alighting information at each stop, transfer information, and more.

Furthermore, public transport movement information can often be collected through Application Programing Interfaces (APIs) in the form of transit combinations provided by portal maps such as Google and Bing.

TABLE 4 Public transport data

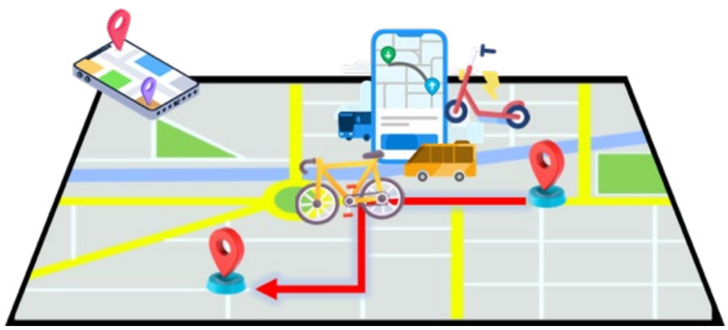


CATEGORY
Mobile Communication Data
Smart Card Data
BIS- and BMS-based information data

Transport modes' trajectory data

In most transport systems, information about individual users and the movement of the modes of transport can be obtained together. GPS-based individual user information collected through smartphone apps includes data on the modes of transport used and travel trajectories. According to the concept of mobility-as-a-service (MaaS), the operation data of modes such as personal mobility (PM), demand responsive transport (DRT), Shared Bike, etc., are important because together, they can recommend an optimal travel route. By utilizing and analyzing MaaS data, including PM and DRT, it is possible to solve issues in public transport and recommend optimized modes of transport to users. Local governments can improve customized mobility services and establish new transport policies to make the lives of citizens more convenient.

TABLE 5 Other transport modes' trajectory data



CATEGORY	COLLECTED INFORMATION
Mobile Communication Data	[Coverage] Entire country // [Information]personal information, periodic location of passengers (latitude, longitude, time), etc.
Mobility Service //App Data*	[Coverage] Service area//[Information] personal information, history of service use (latitude, longitude, time, fare), etc.

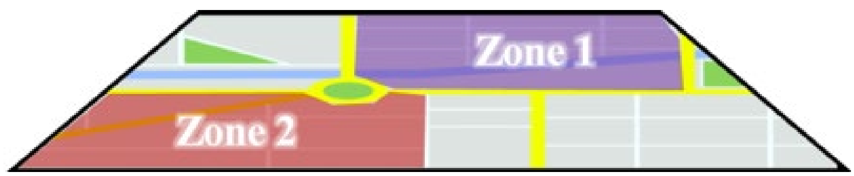
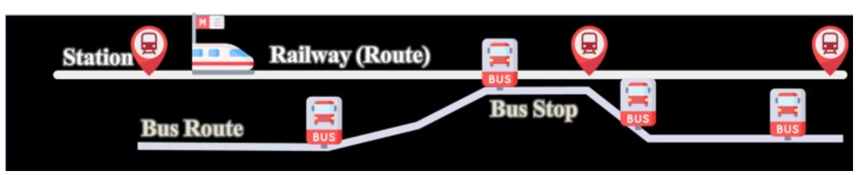
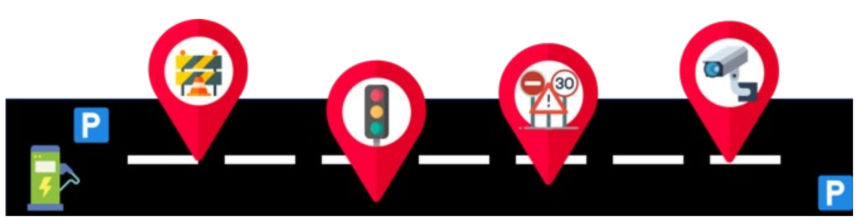
* Mobility Service includes a bike/scooter sharing system, DRT service, etc.

2 Transport and mobility data (operation and facility)

This data usually refers to statistical information that collects a certain location's traffic volume, speed, density, etc., networks (routes, stations, etc.) per modes of private or public transport, and facility data (which are a part of public infrastructure) that are used for traffic operation and maintenance.

Transport and Mobility data is usually managed, stored, and distributed by public institutions, local governments, central administrative agencies, etc., and is also the most basic spatial data for the generation, collection, and analysis of transport and mobility data (activity data).

TABLE 6 Collected information in the operation and facility

CATEGORY	COLLECTED INFORMATION
Road Network Data (Link, Node)	[Range] Entire road network//[Information] Link: Link ID, Link length, Link location information (latitude, longitude) etc., Node: Node ID, Node location information (latitude, longitude), etc.
	
Administrative Districts / Zones (Area)	[Range] National//[Information] Zone ID per Level, Zone name, Zone area, Centroid point (latitude, longitude), etc.
	
Public Transport Network Data (Route, Station)	[Range] Entire region of public transport service operation//[Information] Routes that modes of public transport travel on (railways, BRT routes, etc.): Route ID, name of route, Link ID sets that transfer within a route, route length, route location information (latitude, longitude) / stations that mode of public transport stop at: station ID, name of station, station location information, list of routes that pass through the indicated station, etc.
	
Transport Facility Data (Location Point)	[Range] Entire country//[Information] Facility ID, name of facility, type of facility, facility location information (latitude, longitude), etc.
	

3 Socioeconomic data

Socioeconomic data is a statistic that enables one to make comprehensive and balanced judgments about societal conditions. This data serves various purposes depending on the objective and is managed as indicators that can be maintained over time. The use of socioeconomic data allows for the representation of social values and objectives in a hierarchical form, making it possible to detect social problems early, convey social changes, and intervene in undesirable changes. Additionally, if social reporting becomes institutionalized, social indicators would be able to provide information to the general public about our values and objectives, as well as our current position and direction for the future.

Moreover, international organizations such as the United Nations and World Bank disclose various comparable indicators for measuring living standards and comprehensive social status, predicting social changes, and assessing the performance of social development policies of Asia-Pacific countries. Socioeconomic data provides information about population, labor, housing, transport, environment, welfare, culture, and safety. One of its key features is that it enables a comprehensive measurement of the development of urban and transport sectors, as well as the establishment of various policy plans, decision-making processes, and impact assessments. This is achieved through the integration of Transport and Mobility Data.

4 Global fundamental geospatial data

Urban Spatial Data refers to spatial data provided in the form of digital maps, satellite images, GPS data, and other forms that include various facilities, such as administrative boundaries, bridges, road networks, transport infrastructure, pedestrian walkways, public parking spaces, and clusters of buildings within a city. It used to be defined as an efficient processing technology that surpassed the collection, storage, management, and analysis capabilities of spatial data. Now, it encompasses all data that can be spatially integrated or geospatially enabled, including public sector-accumulated administrative information and public sector Big Data. Urban Fundamental Geospatial Data can generate a variety of values through an integrated analysis with Transport and Mobility Data and also socioeconomic data, supporting policies related to real estate, transport, regional development assessment, and others.

3.2 Case studies of public data portal building per country

The purpose of transport Big Data is to gain credibility regarding the evaluation of various investment businesses for transport infrastructures by commonly utilizing standardized and consistent time series.

Therefore, in this section, questions such as how open operators – national governments, public institutions, private firms, etc. – are being with each category of Big Data (a total of 4 categories) through their use of public data portals, and how useful accessible Big Data is in Big Data-based policy decision-making are reviewed.

Public Big Data, also known as 'Open data,' can be freely accessed, used, modified, and shared by anyone, regardless of their aims. It also enables the timely data collection of a wide range of demographic groups, especially those who are socially and economically underprivileged. The majority of countries make their Big Data and statistical information based on this Big Data accessible to anyone through their Public Data Portal.

The overview and transport Big Data status of the 10 major Asia-pacific countries that possess Open Data Portals (based on the definition of Transport Big Data mentioned in the previous section) is as follows. Countries demonstrate their innovative data utilization capabilities through public data and statistical systems.

In particular, countries that possess stellar Big Data utilization case studies, such as the Republic of Korea, and Singapore, are considered to be at a 'Mature Level' of data integration. It is obvious that they are not only building 'Data Portals' that are web-based and in the form of interactive data and metadata platforms, but also own data analysis tools and perform the role of warehouses.

TABLE 7 Open data portal in Asia and the Pacific

CATEGORY	REGION	DATA PORTAL	CONTENTS				DATA INTEGRATION*
			TRANSPORT AND MOBILITY		SOCIO-ECONOMIC FACTOR	URBANSPATIALDATA	
			OPERATION	MOVEMENT AND TRAJECTORY			
1	**Republic of Korea**	Data.go.kr www.ktdb.go.kr www.bigdata-transportation.kr Data.nsdi.go.kr www.datastore.or.kr	O	O	O	O	Mature
2	**Singapore**	Data.gov.sg www.sgx.com www.nrf.gov.sg datamall.lta.gov.sg www.smartnation.gov.sg apiservices.iras.gov.sg	O	O	O	O	Mature
3	**Australia**	Data.gov.au Aurin.org.au nationalmap.gov.au Ausdatastrategy.pmc.gov.au	O	O	O	O	Developing
4	**Indonesia**	Data.go.id tanahair.indonesia.go.id	O	O	O	O	Developing
5	**Viet Nam**	Open.data.gov.vn Mic.gov.vn	O	X	O	O	Developing
6	**Thailand**	Data.go.th Gbdi.depa.or.th Org.iticfoundation.o	O	X	O	O	Developing
7	**Malaysia**	Data.gov.my www.dosm.gov.my	O	X	O	O	Early-stage
8	**India**	Data.gov.in Indiadataportal.com	O	X	O	O	Early-stage
9	**Bangladesh**	Data.gov.bd	O	X	O	O	Early-stage

Note: *= "Mature" refers to having an 'Open data market' in cooperation with the private sector, "Developing" refers to having plans for data integration with the private sector, "Early-stage" refers to not having enough plans for boosting partnership between the public and private sectors.

3.3 Current status review of Big Data-related regulations

TABLE 8 Big Data-related regulations in Asia-pacific countries

	LOCATION	CURRENT LEGISLATION	REGULATOR	SYSTEMATIZATION
1	**Republic of Korea**	Personal Information Protection Act, 2011	Personal Information Protection Commission http://www.pipc.go.kr/cmt/main/english.do	Mature
2	**Japan**	Act on the Protection of Personal Information, 2017 (APPI) & Amended APPI 2020	Personal Information Protection Commission https://www.ppc.go.jp/en/	Mature
3	**China**	People's Republic of China Cybersecurity Law 2017	Cyberspace Administration of China http://www.cac.gov.cn/	Mature
4	**Singapore**	Personal Data Protection Act, 2012	Personal Data Protection Commission https://www.pdpc.gov.sg/	Mature
5	**Australia**	Privacy Act 1988	Office of the Australian Information Commissioner https://www.oaic.gov.au/	Developing
6	**Indonesia**	• Government Regulation No. 71 of 2019 (GR 71/2019) • Law No. 11 of 2008 on Electronic Information and Transactions (EIT Law) and its amendment Law No. 19 of 2016 • MOCI Regulation No. 20 of 2016 (Protection of Personal data in an Electronic System) (MOCI Law)	Ministry of Communication and Information Technology https://www.kominfo.go.id/	Developing
7	**Viet Nam**	• Government Regulation No. 71 of 2019 (GR 71/2019) • Law No. 11 of 2008 on Electronic Information and Transactions (EIT Law) and its amendment Law No. 19 of 2016 • MOCI Regulation No. 20 of 2016 (Protection of Personal data in an Electronic System) (MOCI Law)	Ministry of Information and Communications https://english.mic.gov.vn/Pages/home.aspx	Developing
8	**Thailand**	The Personal Data Protection Act B.E. 2562 (2019) (PDPA)	Personal Data Protection Committee	Developing
9	**The Philippines**	Data Privacy Act of 2012	National Privacy Commission https://www.privacy.gov.ph/	Developing
10	**New Zealand**	Privacy Act, 1993 & Privacy Act 2020	Privacy Commissioner's Office https://www.privacy.org.nz/	Developing
11	**Lao People's Democratic Republic**	Law on Electronic Data Protection, 2017	Ministry of Posts and Telecommunications https://www.mpt.gov.la/	Developing
12	**Brunei Darussalam**	Data Protection Policy 2014	Minister at the Prime Minister's Office http://www.pmo.gov.bn	Early-stage
13	**Myanmar**	Law Protecting the Privacy and Security of the Citizen, 2017	Ministry of Home Affairs http://www.myanmarmoha.org/ Ministry of Communications and Information Technology http://www.mcit.gov.mm/	Early-stage
14	**Malaysia**	Personal Data Protection Act, 2010	Personal Data Protection Department http://www.pdp.gov.my Malaysian Communications and Multimedia Commission https://www.mcmc.gov.my/	Early-stage
15	**India**	Information Technology Act, 2000	Ministry of Electronics and Information Technology https://meity.gov.in/	Early-stage
16	**Bangladesh**	Digital Security Act 2018	Digital Security Agency http://www.btrc.gov.bd/useful-links/posttelecommunication-division-ministry-posttelecommunications-and-information	Early-stage
17	**Cambodia**	The Constitution of the Kingdom of Cambodia and the Civil Code of Cambodia, 2007 (the Civil Code)	None	Early-stage
18	**Mongolia**	Personal Secrecy (Privacy) Act, 1995	None	Early-stage
19	**Papua New Guinea**	Privacy Communications Act, 1973, Constitution of the Independent State of Papua New Guinea and National Information and Communications Tech Act, 2000	None	Early-stage
20	**Sri Lanka**	N.A.	Information and Communication Technology Agency https://www.icta.lk/	Early-stage

Before a country can perform Big Data-based policy decision-making, it needs first to have an environment for the implementation of Big Data-related policies, such as Big Data utilization and information protection. In terms of information protection, there needs to be regulations regarding the responsibility of personal information leakages, data information security, measures against data discrimination, and the protection of data as public or private assets. Furthermore, there should be a solution that allows the country to overcome the barriers to Big Data utilization. For example, mobile base station data, vehicle GPS data, and Smart Card data all include personal information, meaning that there are inevitable limitations to sharing them and incorporating them in policies.

Many countries in the Asia-Pacific region are working to improve data connectivity, openness, and quality for the enhancement of Big Data usability. After reviewing the current status of Big Data-related legal policies, it was observed that developed countries and countries with developing levels at the systemization level or early-stage level have high interests in the protection of personal information and data information security.

Along with the Personal Data Protection Act, countries are also strengthening their legal policies for the enhancement of Big Data usability by enacting public data laws, data privacy protection laws, etc. However, countries lack the understanding of social values and data-based administration, often treating data as a particular institution's property, resulting in the inability to establish a systematic policy related to Big Data. This predicament is likely to limit the usability and decision-making process of data as a result of problems such as data procurement issues due to the selfishness of public offices, the issue of being unable to consider the form of demand due to a supplier-centered statistical data collection, data privatization, and more.

Hence, Big Data policies are aiming to improve the usability of public data through private firms. This is done by prioritizing the further opening of public data by creating an organized environment that enables the maximization of data usability. As a concept that includes both public and private data, Big Data policies should strengthen the connection between public and private data and enhance its usability. In most developed countries such as the United States, the United Kingdom, Japan, etc., the development of Big Data policies is focused on the rectification of promotion policies that involve disclosing public data, expanding public services, forming a group of Big Data experts, and personal information protection laws.

In order to use Big Data for each country's official statistics, policies need to be strengthened, and appropriate technologies need to be developed. This process requires soft skills and credibility as well as time and patience. Thus, the National Statistical Office needs to develop a new business model to take advantage of labor potential, decision-making abilities, and data resources, and could also enact and amend laws that can back this up. Through this, the connection with policy frameworks, private-public partnerships, and various data science organizations needs to be strengthened.

National data collection institutions should promote official statistics as a public good and should consistently improve the data's quality, such as its timeliness, categorization, and significance. Countries will need guidelines such as Big Data-related statistical standards and factual resources. Also, the National Statistics Office could approach the Big Data they possess for development purposes and could continuously prepare legal procedures to prevent Big Data being used in inappropriate ways.

Finally, in order to improve transport Big Data utilization, the issues of clarifying data ownership and the standardization of data prices could be considered. Although data is currently being traded by bargains between supplying institutions and demanding institutions, in the future, there has to be an environment where various types of Big Data can be freely shared through a government-led standardization of trade. As it is difficult to build an integrated database in the private sector (where the ownership of data is scattered), there should be government-led data integration, building, and sharing, as well as the proposal of relevant policies.

3.4 Investigation of transport Big Data analyses and case studies

Transport Big Data analyses and case studies are divided into a total of 4 categories: Transport Planning, Transport Monitoring and Operation, Public Transport, and Transport Safety.

Each case study has been classified to be at a strategic level, tactical level, or operational level depending on their planning level. Cases at a strategic level refer to cases that collect a city's data through various Big Data collection systems and evaluate long-term traffic networks or produce solutions for further improvement. Cases at a tactical level gather data through a single or several Big Data collection systems and have used such data for the design and enhancement of a certain mode's route or network. Finally, cases at an operational level provide direct transport services through Big Data collection systems that enable real-time or short-term collection or utilization.

TABLE 9 Big Data-based transport sector analysis and case study types

CATEGORY	CATEGORY	CATEGORY
Transport Planning	PUMA (World Bank)	Transport Planning
	AURIN (Australia)	Strategic Level
Transport Monitoring & Operation	Transport Monitoring & Operation	Transport Monitoring & Operation
	City Brain (China)	Operational Level
	Smart Nation (Singapore)	Strategic Level
	FASTER (Singapore)	Operational Level
Public Transport	Public Transport	Public Transport
	Night Bus (Republic of Korea)	Tactical Level
	Beijing Transport Big Data Project (China)	Strategic Level
Transport Safety	Transport Safety	Transport Safety
	High-Risk Road Forecasting//System (Republic of Korea)	Operational Level

Source: ESCAP , Increasing the Use of Smart Mobility Approaches to Improve Traffic Conditions in Urban Areas of South-East Asia (Accessed on August 18, 2023).

1 Transport planning

A PUMA (Platform for Urban Management and Analysis, World Bank)

The World Bank has created a Big Data-based evaluation platform called the Platform for Urban Management and Analysis (PUMA), which provides socioeconomic data for East Asia and the Pacific, including information about built-up land, urban expansion, and urban population changes from 2000 to 2010. This platform enables a comparison of urbanization levels between countries that were previously defined in different ways. In addition to containing data on land use, city expansion, and population, PUMA has information about transport networks (see Table 10). This allows PUMA to be used for comprehensive city analyses and the evaluation of various transport systems, including infrastructure investments that can overcome the limitations of using surveys or projected data.

TABLE 10 List of data from PUMA

THEME	VISUALIZATION
Land cover	Land cover types/classes
	Share of artificial land
	Structure of artificial land
	Intensity of land cover change – total by flow
	Net formation of land cover by class
Population	Population development
	Population: distribution and density
Transport	Density of road network
	Average road width
	Road length by width category
Urban expansion	Urban expansion
	Total area of built-up land
	Share of built-up land (% of total area)
Population	Distribution and density of population in built-up areas
	Change of population in built-up areas
Total population	Distribution and density of total population

B AURIN (Australian Urban Research Infrastructure Network)

The Australian Urban Research Infrastructure Network (AURIN) is an initiative under Australia's Government National Collaborate Research Infrastructure Strategy (NCRIS), which aims to use research to understand how cities and urban areas function as complex systems. AURIN aims to analyze the behavior, production, and consumption of people within definable demographic, social, economic, and environmental parameters.

AURIN leads a collaboration between researchers, planners, and policymakers in Australia to develop the Australian Transport Research Cloud (ATRC). The ATRC platform, funded by the Australian Research Data Commons (ARDC), delivers streamlined data and tools to accelerate transport research and changes across Australia. AURIN's platform manages and analyzes Big Data of urban areas and transport to resolve issues related to pollution, safety, health, and unemployment. The collaboration includes academia and the public and private sectors and contains over 3,500 data sets from 98 entities. These include urban development and transport data, such as property ownership and characteristics, street maps, parcels, census tract data, assessed values, and land use.

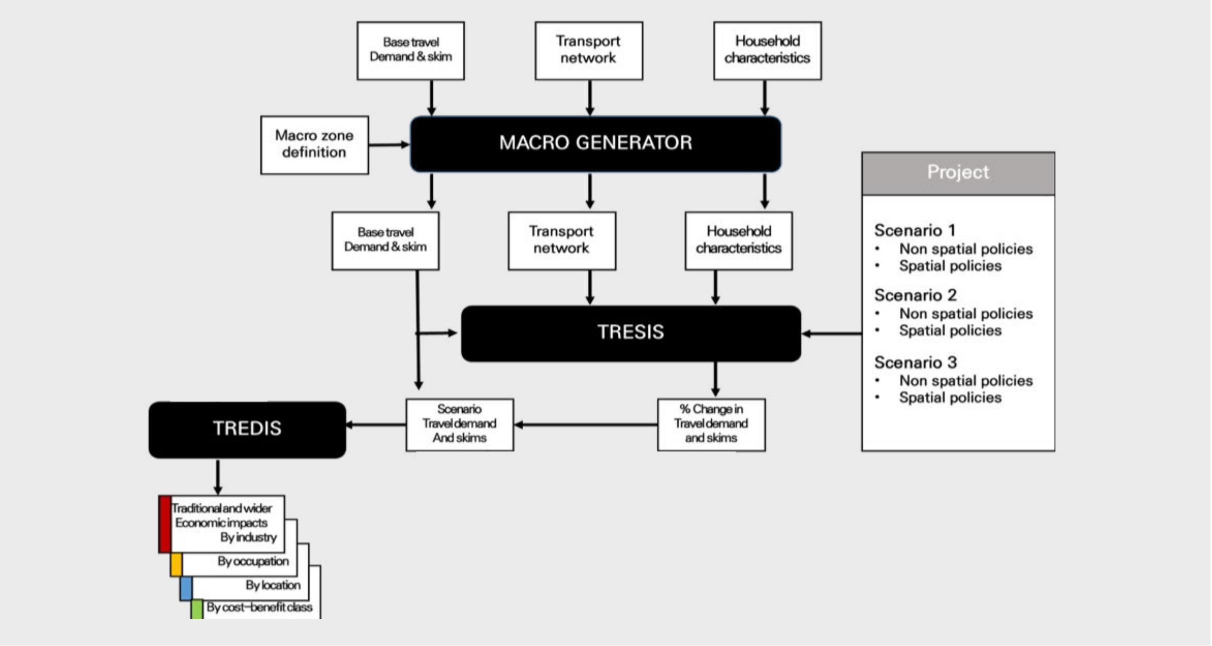
AURIN has provided Australian researchers with capabilities to develop an evidence-based policy for informed and effective decision-making to shape communities and urban environments. The ATRC platform extends AURIN's existing infrastructure to enable a better understanding of current and future supply and demand of transport and mobility services in Australian cities and towns. It facilitates collaboration between government, research, and industry to enhance transport services and infrastructure planning. ATRC's objectives include making public transport more convenient, accessible, and efficient; reducing road congestion and vehicle emissions; improving urban mobility and reducing commute times; improving national productivity and international competitiveness; reducing air pollution, noise pollution, and urban heat islands; and improving population health and well-being.

FIGURE 3 AURIN platform

Source: AURIN, Available at https://aurin.org.au/.

MetroScan is also an example of an economic impact analysis tool that uses AURIN's massive database to evaluate and implement investment in transport infrastructure. As shown in Figure 3 MetroScan goes beyond merely calculating the accessibility of transport investment or mobility opportunities by taking into account the additional productivity and social values generated within the local economy. By utilizing Big Data, this tool is able to improve its predictive abilities, making it possible to evaluate the feasibility and effectiveness of new transport infrastructure investments with greater accuracy.

FIGURE 4 MetroScan framework



Source: AURIN, Available at https://aurin.org.au/.

2 Transport Monitoring & Operation

A Transport for London WiFi Project (UK)

In the Transport for London WiFi Project, Transport for London (TFL) partnered with a private company to provide free Wi-Fi to 97 per cent of London's subway stations. TFL then collected Wi-Fi usage data, which was combined with preexisting oyster card and non-contact payment card data to create a comprehensive picture of users' transport habits. This includes travel time between stations, travel routes, and maximum waiting times. This data can be used to improve subway schedules, station designs, and major stations in order to create a better transport system.

By analyzing additional data, such as changes in congestion, travel patterns, and waiting times, TFL can provide real-time information to help users avoid congestion when planning their travels. The system also enables efficient management of service suspensions and other unique circumstances to enhance information delivery and services for users. Additionally, the Wi-Fi data can be used to understand user behavior and maximize profits from subway advertisements and retail stores.

It is important to note that TFL only collects travel pattern data and doesn't collect any personal information. Users who do not wish to have their travel pattern data collected through Wi-Fi can simply turn off their Wi-Fi settings or device, or switch to airplane mode to avoid data collection.

FIGURE 5 Wi-Fi data collection in TFL

Source: Available at https://tfl.gov.uk/corporate/publications-and-reports/wifi-data-collection.

B Smart Nation Project (Singapore)[1]

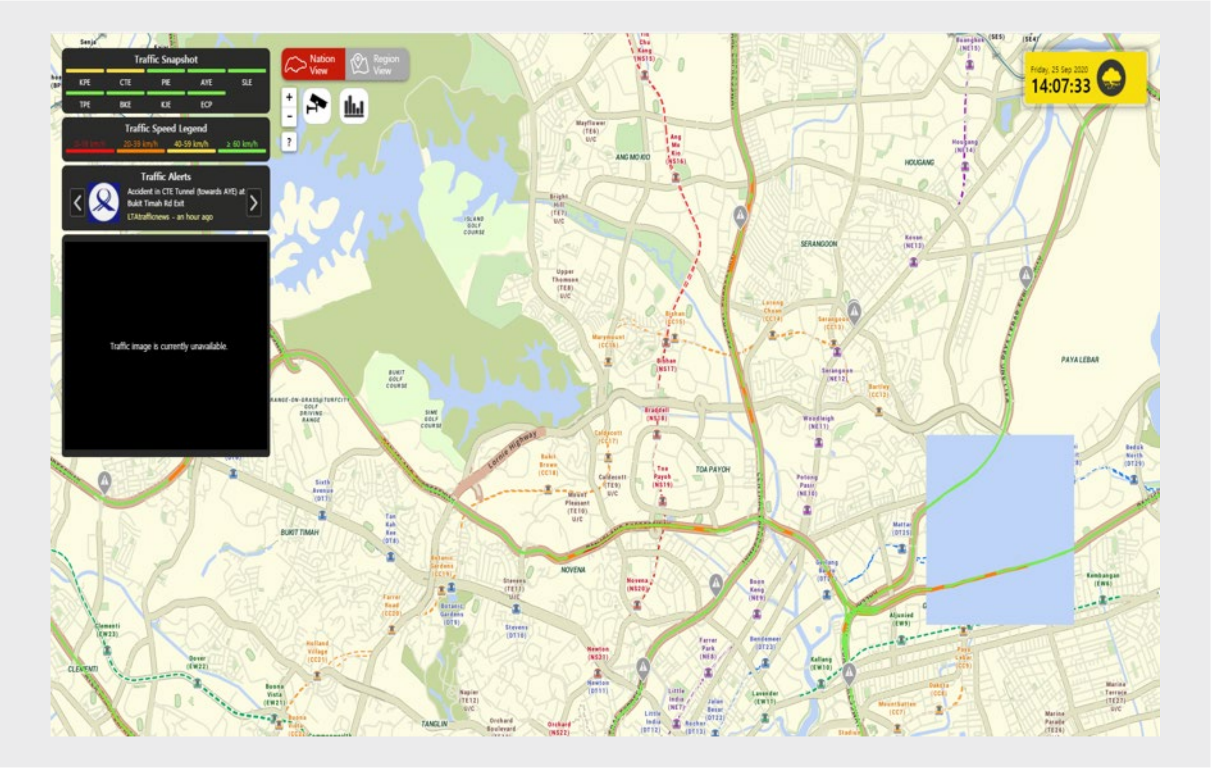
The Smart Nation Project was initiated in 2014 by the Prime Minister of Singapore to transform Singapore's economy into a digital economy through technology. The project focuses on various areas within the transport sector, including autonomous vehicles, contactless fare payments, on-demand shuttles, open data and analytics for urban transport, and research to develop standards for autonomous vehicles.

One of the themes under open data and analytics for urban transport is the use of Big Data from transport card systems to improve public transport services. This involves identifying congested areas during peak commuting hours using data analytics. Additionally, real-time location data and arrival time data from sensors installed in over 5,000 vehicles are collected to make public transport plans that better meet

1 Smart Nation Singapore, "Homepage". Available at https://www.smartnation.gov.sg/ (Accessed on August 18, 2023).

commuters' needs. This has resulted in an average reduction of three to seven minutes in bus waiting times. Furthermore, the data is utilized to understand traffic congestion and patterns in real-time by linking it with other data acquired from other Smart Nation projects.

FIGURE 6 Real-time Singapore traffic watch



Source: Smart Nation Singapore, https://www.smartnation.gov.sg/initiatives/transport/open-data-analytics.

C City Brain (China): An Urban Traffic Management System[2]

In 2016, a collaboration between the private sector and the city of Hangzhou, China, led to the development of a new project called "City Brain." This project aimed to construct an urban traffic management system to improve road conditions and monitor incidents in downtown areas. From 2017 to 2018, the City Brain 1.0 project focused on controlling traffic signal timings from 128 CCTVs based on traffic volumes and vehicle movements. Various entities, including transport ministries, traffic control, safety, planning departments, and service providers shared and exchanged diverse data types. Data mining techniques were applied to identify road congestion levels and provide real-time traffic information to users via smartphones. The City Brain 1.0 project generated 250 million data transactions daily, making it possible to efficiently manage the urban transport networks without any human effort. City Brain 2.0 is now in the planning stages and is expected to improve public services for citizens, including city security, in addition to the smart transport systems.

2 Sohu, "Hangzhou has "wisdom" in tackling congestion (big data observation)", Available at https://www.sohu.com/a/126476791_114731 (Accessed on August 18, 2023).

FIGURE 7 City Detection and Analysis

Source: WIRED, "In China, Alibaba's data-hungry AI is controlling (and watching) cities", Available at https://www.wired.co.uk/article/alibaba-city-brain-artificial-intelligence-china-kuala-lumpur.

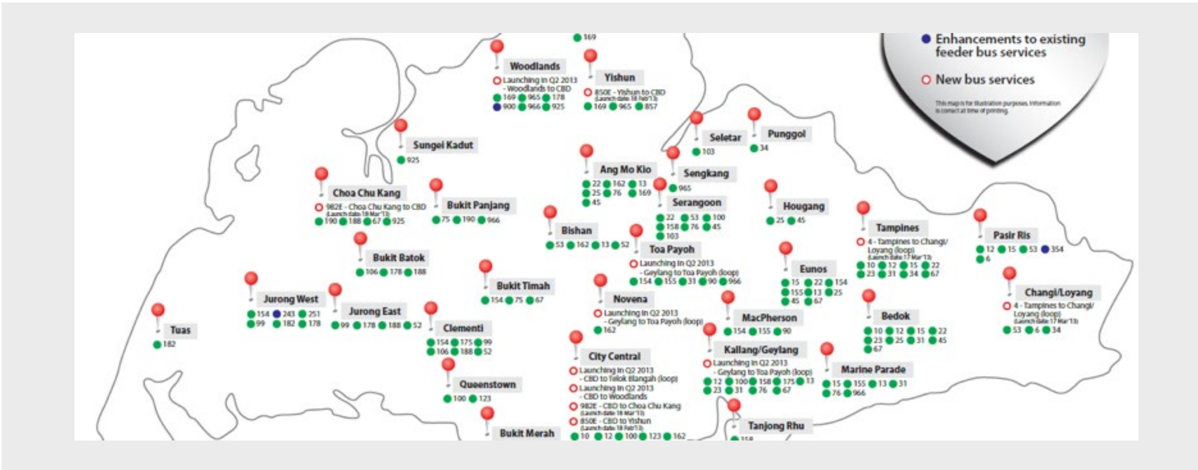
3 Public transport

A Bus Service Enhancement Program (BSEP, Singapore)[3]

The Bus Service Enhancement Program (BSEP) was initiated by Singapore's Land Transport Authority with two objectives: the purchase of new buses and the improvement of peak hour travel. Between 2012 and 2017, the program purchased 1,000 new buses and expanded 218 bus routes, including feeder and short trunk services, to developing areas. To achieve the objectives of BSEP, Big Data was used by analyzing bus management data on the service provider side and Smart Card data on the service user side. This data was used to identify regions where the supply of bus services did not meet the demand. The analysis supported the addition of routes which directly connected these regions to downtown areas and introduced new shuttle services during peak hours to equalize demand and supply. BSEP also employs navigation systems to optimize intervals and waiting times and to readjust bus service lines in real-time. By using Big Data, BSEP can assess bus service lines and bus stop locations, considering the service provider's and users' needs. One advantage of using Big Data is its ability to meet users' demands in real-time to shorten travel times and improve location services for users.

3 Land Transport Guru, "Bus Service Enhancement Programme (BSEP)", Available at https://landtransportguru.net/bus-service-enhancement-programme-bsep/ (Accessed on August 18, 2023).

FIGURE 8 Bus Service Enhancement Program

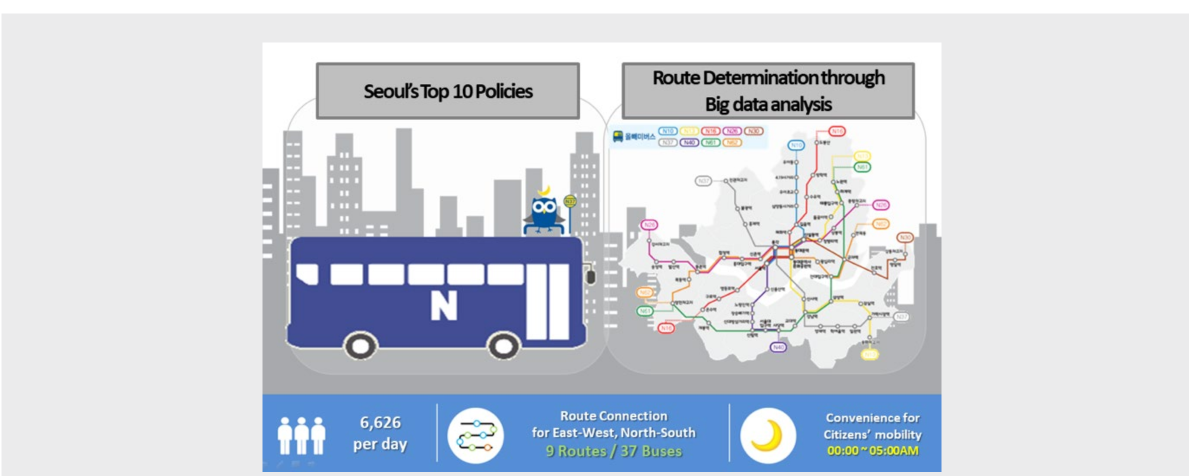


Source: Available at https://www.facebook.com/WeKeepYourWorldMoving/photos/a.137686272968284/410022175734691/?type=3.

B Seoul night bus[4]

A late-night bus service, called Owl Bus has been in operation since 2012 to supplement the shortage of late-night transport services in Seoul. Public transport services are generally inadequate late at night because of low economic activity. Also, taxis and chauffeur services are burdensome because they are relatively expensive, particularly for low-income groups. The Owl Bus project started to address such issues and used Big Data to establish bus service plans. A variety of data was incorporated into this project, including mobile phone usage data and smart card data from taxis to identify pick-up and drop-off destinations for late-night transport users. The data was utilized by dividing the entire area of Seoul into cells, where a foot traffic and their travel demands were recognized, and then used to analyze existing bus service lines and times by each date in order to identify the optimal service lines and service frequency. A radial transport network was constructed to connect nine areas on the outskirts of the city to make user journeys more convenient. As of now, 70 Owl Buses are being operated on nine service routes from 11:30 p.m. to 6:00 a.m.

FIGURE 9 Owl Bus Concept

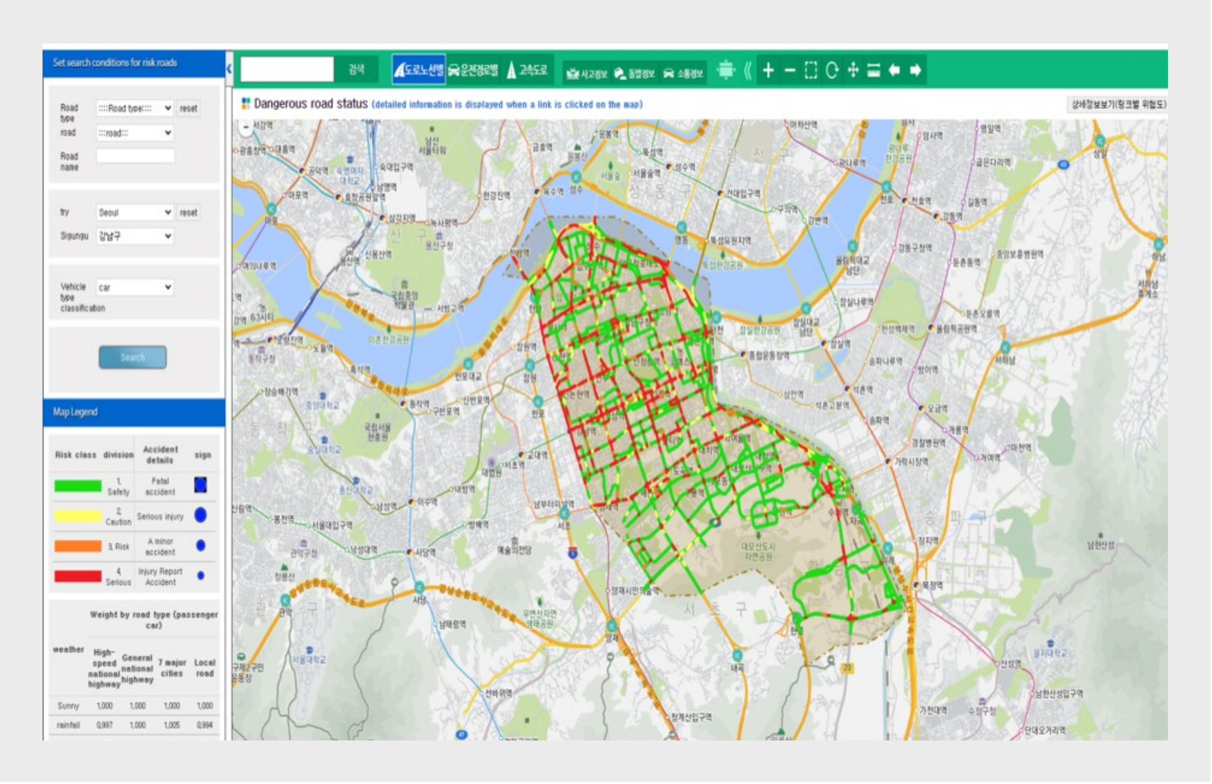


4 Available at https://www.seoulsolution.kr/en/content/night-bus-called-owl-bus-route-design-using-big-data.

4 Transport safety

In the Republic of Korea, the Traffic Accident Analysis System (TAAS) has provided a range of statistics about traffic accidents and traffic safety information since 2008. The objective of TAAS is to offer a high level of traffic safety across all roads in the country. Big Data is also used to predict the degree of real-time or near-future crash risks on the roads. Due to the fact that traffic accidents occur because of many reasons, such as human-, vehicle-, and road environment-related factors, TAAS employs various types of information, such as previous accident records, weather conditions, unexpected incidents, and real-time traffic conditions. Four levels (safety, caution, risky, and serious) of safety are presented by this system based on correlation analysis among factors and the prediction of the degree of road risks for traffic accidents.

FIGURE 10 Traffic Accident Analysis System



Source: KoRoad, "TAAS", Available at http://taas.koroad.or.kr/TCFS/ 9 (Accessed on August 18, 2023).

Chapter 4

Alternative solutions for the promotion of transport Big Data-based policies

4.1 Understanding Big Data-based policy decision-making

What is data-based policy decision-making?

In order to define the concept of 'data-based policy decision-making,' the origin and meaning of the phrase, 'data-based,' as well as the process of policy decision-making should be understood.

The origin of 'data-based' is the phrase 'evidence-based' used in health and medical fields. This concept is based on the importance of incorporating factual information in the careful decision-making process of diagnosing a patient. It was introduced to policymaking in March 1999, as the U.K. government proposed the "White Paper: Modernizing Government" project. The United States, on the other hand, presented "The Obama evidence-based social policy initiative" under Obama's first and second administrations, established the "Commission on Evidence-Based Policymaking a.k.a. CEP" in 2017, enacted the "Foundations for Evidence-Based Policymaking Act" in 2018, and is now making an effort to generate, manage, analyze, and activate evidence for policy decisions in the public sector. These countries define the quality of usable information by limiting it to 'information produced by "statistical activities" with a "statistical purpose" that is potentially useful when evaluating government programs and policies'. In this sense, it would be more fitting to say 'data-based' or 'statistics-based' rather than 'evidence.'

Necessity of data-based decision-making

Data-based decision-making is important because it is based on actual data rather than past experiences or opinions. It is especially crucial in the transport sector, which is closely related to the lives of citizens and is capable of collecting a large amount of data through various sources. With a detailed data analysis, it becomes possible to identify regions with high volumes of traffic and accident rates, which thereby increases the overall efficiency of the transport system. In addition, it suggests a guideline on how to divide a limited budget and enables the measurement and evaluation of the impacts of an established policy.

Big Data analysis, which uses various forms of data, is being carried out in many fields, especially those related to resolving regional issues and making policy suggestions. There needs to be a platform or infrastructure that collects and stores data in order to inspect a region's status using data and to provide appropriate solutions for such issues. A massive volume of data collected and stored in this manner goes through processing and fusion procedures suitable for various subjects and is used in policy decision-making or to predict regional issues. Such data becomes especially useful for data-based policy decision-making.

Data-based decision-making can help immensely when policies are rewritten or edited in the decision-making process. The table below shows the procedures basic data goes through in order to become meaningful data for policy decision-making. In the beginning, data is processed, arranged, and then loaded onto a database in the form of 'Information.' The loaded information is used in analysis to solve various social issues, and these analysis results are used as 'Evidence' for making policy suggestions.

'Knowledge' refers to remedial knowledge about various issues that can occur in a certain region, rather than a specific issue. This knowledge is constructed after considering the region's temporal and spatial characteristics as well as its social conditions. 'Wisdom' is a combination of 'Evidence' and 'Knowledge.' It represents the policy direction that is set by the central or local government after considering the reactions of the general public and the region's characteristics. 'Wisdom' is the most advanced and difficult level of data-based decision-making and refers to making a strategic decision while considering numerous factors and circumstances in order to solve a complex problem.

TABLE 11 Hierarchy of Data based decision making

CATEGORY	HIERARCHY	SHAREABILITY	CONTENTS
Wisdom	↑	Difficult	Central and local governments set a policy direction after considering various public opinions
Knowledge	↕	↑	Remedial knowledge that reflects the regional society's temporal and spatial conditions and characteristics
Evidence		↕	Topical information sets or analysis results of scenarios for resolving pending issues
Information		↓	A databased collected and stored from basic data
Data	↓	Easy	Basic temporal and spatial data

4.2 Big Data-based policy decision-making stage

In the Republic of Korea, a law addressing data-based administration activation was enacted in 2020. This law defined data as 'information that has been generated or processed through a machine which has information processing abilities, and all structured or unstructured information that exists in a form that can be deciphered by a machine'[5]. This is a definition that is slightly distinct from the original context of 'evidence-based,' as it considers the utilization of unstructured data in the modern industry. However, the importance of informatization and standardization applies to the government's utilization of structured Big Data as well, such as in administrative databases and financial business management systems. Ultimately, 'data' in 'data-based' refers to information that has been generated through systematic methods and statistical activities while existing in a form that can be deciphered by machines.

Upon reviewing cases of countries that have proceeded with Big Data-based policy decision-making, the majority of cases proceeded under the purpose of further investigating the national and regional social issues and the reasons behind the poor performance of policies. In addition, a country's implementation stages of data-based policy decision-making (which are carried out in the process of investigating solutions to problems) can be divided into four steps: data, insight, decision, and action. The divisions within the hierarchy of data-based decision-making (defined above) are reviewed by policymakers as essential components of each step.

TABLE 12 Steps of data-based decision-making

STEP 1	STEP 2	STEP 3	STEP 4
DATA	INSIGHT	DECISION	ACTION
	←	→	
	With Information, Evidence	With Knowledge	With Wisdom
			

5 Article 2, Paragraph 1: Korean Law addressing data-based administrative action.

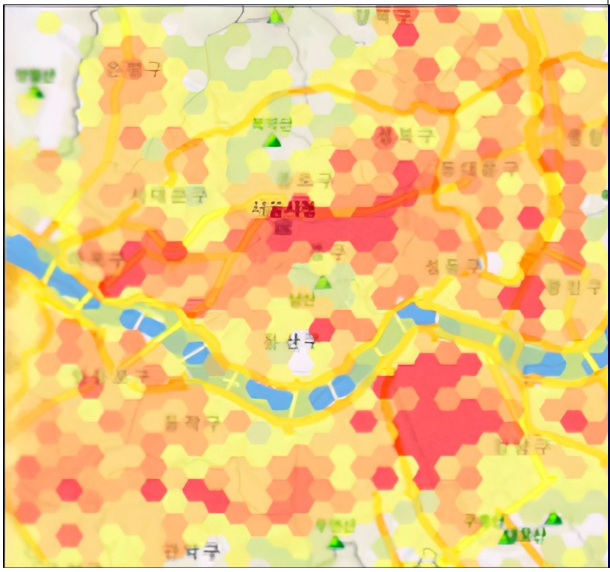
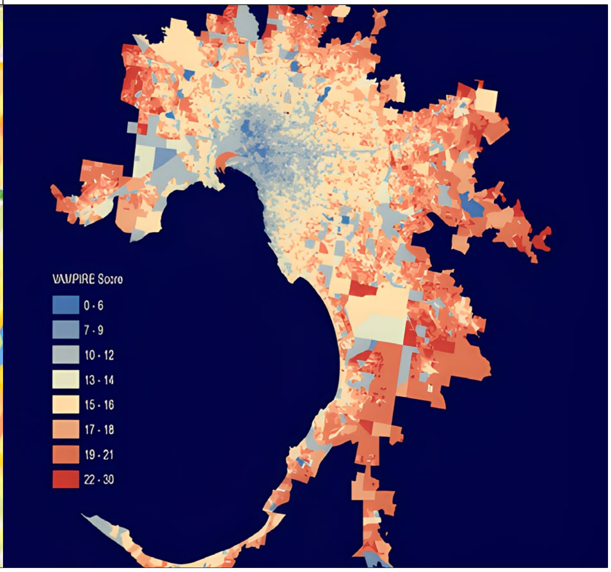
1 Data

Transport sector data refers to transport-related data that has been built through the standardized investigation and analysis of basic information about transport facilities (roads, railways, airports, harbors, distribution facilities, etc.), the operational conditions of transport modes, traffic volume, travel characteristics, and traffic networks. To facilitate the comparison between countries, this report reviewed how much transport data Asia-pacific countries possessed in the previous chapter, mainly focusing on public data that can be used in data-driven decision-making.

2 Insight

Insight is the step where hypotheses are verified, and current situations are diagnosed through the results of data analysis. To do this, statistical testing techniques and transport-specific indicators are used to explain phenomena, and statistical visualization techniques and mapping methods are used for spatial data. Insight encompasses all stages of recognizing problems, allowing decision-makers to review current factors before making final decisions.

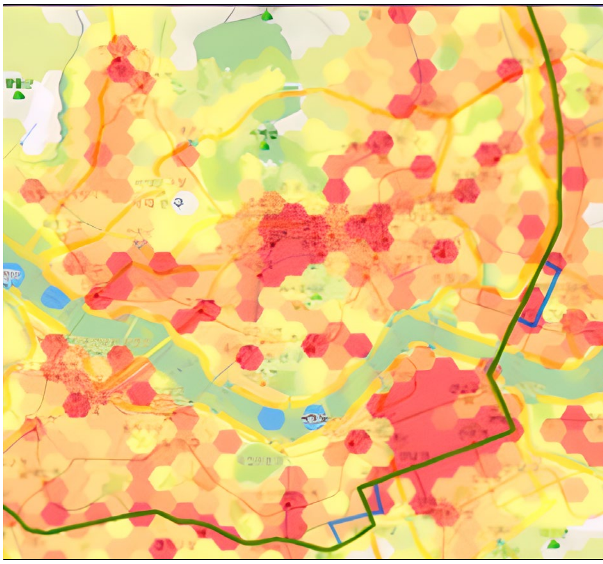
TABLE 13 Comparison of two key cases in the view of insight

CASE 1 (SEOUL)	CASE 2 (AUSTRALIA)
(Night bus) Visualization of the concentrated area of the floating population	Map of Melbourne showing the VAMPIRE score (the Vulnerability Analysis for Mortgage, Petroleum and Inflation Risks and Expenditure)
	

3 Decision

In the Decision step, insights are drawn using data, an in-depth evaluation of diverse solutions (route optimization, network analysis, etc.) is conducted through various algorithms and simulations, and possible solutions are looked into. Amongst these solutions, the most optimal solution for the Action step can be assessed.

TABLE 14 Comparison of two key cases in the view of decision

CASE 1 (SEOUL)	CASE 2 (AUSTRALIA)
(Night bus) Bus route optimization with floating population	Network Analysis using Scenario
	

4 Action

The Action step is where the most optimal alternative is selected among the available solutions. For instance, Republic of Korea had analyzed the demand of nighttime buses and added more bus lines. In this final step, an in-depth review of the feasibility of alternative routes is conducted and a final decision is made. As another example, the Singaporean government is currently being aided by the National Research Foundation to build Virtual Singapore.

Up until this point, the four steps needed for the implementation of transport data-based policy decision-making were examined. Through these steps, Big Data technology portfolios provided safer, cleaner, and more efficient transport services and enabled the management of data necessary for solutions regarding user-customizing issues. The recent transformation of transport Big Data provided new opportunities and services by using new data sources and techniques. This change is seen in existing problems (real-time traffic monitoring) and also through the facilitation of mobility pattern recognition using new data sources such as mobile data.

This background has facilitated the four-step structure of decision-making by enhancing and diversifying data sources (step 1) to discover new insights (step 2), reviewing these insights to search for possible alternatives in the decision step (step 3), and finally translating these alternatives into actual Actions (step 4). Big Data uses such information to resolve transport issues of unprecedented scale. If each country's government follows this direction and utilizes spatial Big Data (geospatial context) befitting their conditions, the implementation of new policies and research (for the enhancement of mobility and the transport sector) as well as the development of government platforms are to be expected.

TABLE 15 Comparison of key cases in the view of action

<table>
<tr><th colspan="2">CASE 1 (SEOUL)</th><th colspan="2">CASE 2 (SINGAPORE)</th></tr>
<tr><th colspan="2">(NIGHT BUS IN SEOUL) FINAL ROUTE PLAN</th><th colspan="2">VIRTUAL SINGAPORE</th></tr>
<tr><td colspan="2">
</td><td colspan="2">
</td></tr>
<tr><th colspan="4">CASE 3 (AUSTRALIA)</th></tr>
<tr><td>Data
The ATRC will provide access to high-value reference datasets to support informed, world-leading transport research, including transport network and household travel survey (HTS) data.</td><td>Tools
Researchers will be able to use the ATRC to conduct detailed network analyses, such as assessing accessiblity, and mobility simulations to test current and future scenarios.</td><td>Integration
The ATRC will connect with data and code from various sources, facilitating the development of seamless workflows.</td><td>Sharing
Collaboration will be a core part of the ATRC, with the ability to make research outputs available to other users to encourage community-building that builds on previous research.</td></tr>
</table>

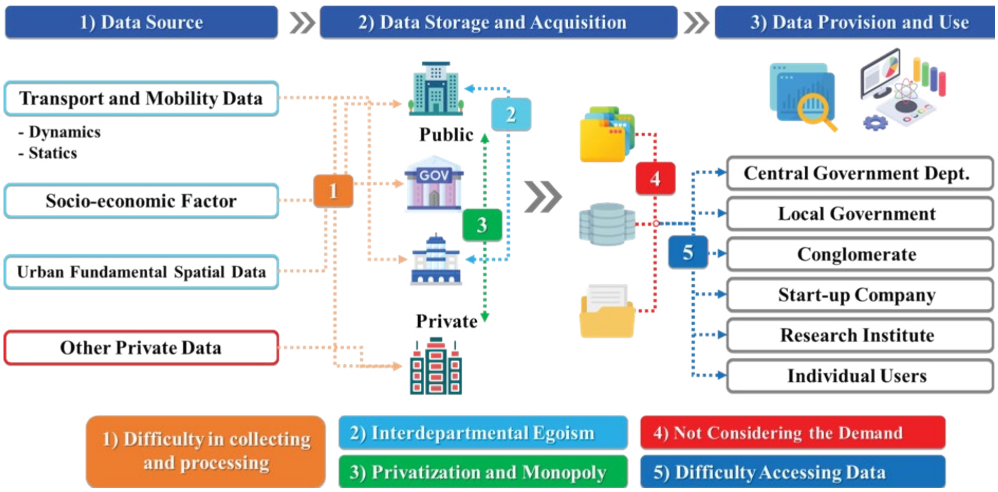
4.3 Limitations of Big Data-based policy decision-making

Although the transport sector possesses a high volume of information that is useful for both the public and private sectors, overall, it tends to lack the systematic organization that enables Big Data-based decision-making. It is therefore essential to create the country's internal and external data utilization environment by thoroughly reviewing the limitations of data collection and individual laws that prevent the maximization of Big Data usability.

The issues and limitations of Big Data utilization can typically be found in these three steps: (1) Data Source, (2) Data Storage and Acquisition, and (3) Data Provision and Use. In particular, 5 limitations that exist along the 3 steps are presented as commonly experienced issues in the public sector. For the advancement of each country's Big Data-based policy decision-making, the limitations that exist in each step should be carefully examined and the levels of utilization should be evaluated. Hence, in this section, the relevant national organizations per limitation (divided into 5 categories) are laid out, as well as cases that each department would have to independently evaluate.

The self-evaluation table below contains 3 Levels. Level 1 indicates 'no understanding or lack of preparation' regarding the evaluation category. Level 2 indicates 'moderate understanding but difficulties in appropriate techniques, financial affairs, and labor utilization at the preparation stage.' Level 3 indicates 'high understanding and capable of strategy establishment.' The evaluation categories were written based on the 'Improving Agency Data Skills Playbook' by the FDS (Federal Data Strategy) of the United States.

FIGURE 11 Limitations and problems of Big Data utilization



The first limitation is 'difficulty in collecting and processing,' which refers to limits that exist in the collection and processing stages of data that make up the foundation of Big Data usage. This limitation exists due to the lack of manpower and wide disparities in data utilization capacities between organizations, leading to excessive time consumption and high costs for the collection and processing of data. If a country is experiencing such issues, a more detailed self-evaluation of the country's (1) level of data understanding, (2) level of data collection limits, and (3) level of data processing limits should be conducted using the table below.

TABLE 16 Self-evaluation in 'difficulty in collecting and processing'

LEVEL OF DATA UNDERSTANDING			
EVALUATION CATEGORIES AND EVALUATION RESULTS (3 LEVELS)	LEVEL 1 (WEAK)	LEVEL 2 (MODERATE)	LEVEL 3 (EXCELLENT)
Understanding of data characteristics, structure, and metadata			
Understanding of storage methods of structured and unstructured data			
Definition of data needed for analysis			
Level of Data Collection			
EVALUATION CATEGORIES AND EVALUATION RESULTS (3 LEVELS)	LEVEL 1 (WEAK)	LEVEL 2 (MODERATE)	LEVEL 3 (EXCELLENT)
Data extraction, conversion, and possession of ETL (Extract, Transform and Load) technology			
Establishment of data storage structure design environment			
Establishment of real-time data collection methods			
Establishment of data connection and supply methods			
LEVEL OF DATA PROCESSING			
EVALUATION CATEGORIES AND EVALUATION RESULTS (3 LEVELS)	LEVEL 1 (WEAK)	LEVEL 2 (MODERATE)	LEVEL 3 (EXCELLENT)
Possession of high-volume data processing technologies and sufficient labor			
Possession of anonymization and de-identification technologies (of personal information) according to personal privacy laws			
Existence of data error (outliers and missing values) searching and editing processes			

The second limitation is 'interdepartmental egoism' and refers to the organizational and structural limits that exist between public institutions that store and manage Big Data. This egoism complicates the data-sharing process between institutions that supply Big Data in the public sector. In the public sector, if an extreme amount of time and effort is consumed due to excessive processes, the motivation to use data can decline. Therefore, efforts to minimize administrative procedures between public institutions, overthrow egoism, and improve communication are the most crucial components of strengthening data utilization capacity. If a country is experiencing such limitations, detailed evaluation of the country's current level of data-centered organization management should be conducted using the table below.

TABLE 17 Level of Data-centered Organization Management

EVALUATION CATEGORIES AND EVALUATION RESULTS (3 LEVELS)	LEVEL 1 (WEAK)	LEVEL 2 (MODERATE)	LEVEL 3 (EXCELLENT)
Existence of a policy guideline regarding data management between relevant institutions under the same sector (ex: transport sector)			
Creation and operation of an organization responsible for the integrated management of data between public institutions or within an individual institution			
Minimization of requested procedures between institutions and departments for data sharing			
Level of communication and teamwork for the smooth sharing of data			

The third limitation is 'privatization and monopoly' and arises when there is a lack of data openness between the public and private sectors during the establishment of a data collection system. The background of this limitation also encompasses the hardships experienced from providing public data in raw form, limiting insight and value investigations needed by the private sector. Thus, from the perspective of the private sector, even if they supply their data to the public sector, if there is neither an adequate integrated connection framework nor a proper compensation system established by the government, the privatization and monopoly of data are inevitable. Additionally, from the perspective of the public sector, if they don't integrate and connect their data to private data, they would not be able to reach the Decision stage, which is necessary for the application of data in policymaking. If a country is experiencing such limitations, a more detailed evaluation of the country's current level of data-centered organization management should be conducted using the table below.

TABLE 18 Level of public/private core data distribution and utilization

EVALUATION CATEGORIES AND EVALUATION RESULTS (3 LEVELS)	LEVEL 1 (WEAK)	LEVEL 2 (MODERATE)	LEVEL 3 (EXCELLENT)
Existence and possibility of future reformation of legal policies about the utilization of personal information			
Existence and introduction of policies for the integrated connection of public/private data			
Existence of an institution/department responsible for the hub & spoke role and overall management of public/private data			
Existence of a potential strengthening program through technology and human resources exchanges between private data firms			
Experience with and future plans for public/private national R&D research			
Existence of a private data transaction and compensation system			

While the fourth limitation, 'not considering the demand' has no difficulties in the aspect of data collection and utilization, it involves the lack of consideration regarding what consumers want, including institutions responsible for policy decision-making, start-ups, research institutions, individual users, etc., which poses a challenge to the direct utilization of data. The fifth limitation is 'difficulty accessing data' and also refers to cases where data is supplied in a manner that is inaccessible or not easily usable by consumers.

These two limitations involve the issue of supplying data under a passive and supplier-centered system. For example, a number of Asia-Pacific countries still fail to consider various consumers and only supply data with fixed forms or statistical data. If a country is experiencing such limitations, a more detailed evaluation of the country's current level of data-centered organization management should be conducted using the table below.

TABLE 19 Level of consideration for consumers of publicly supplied data

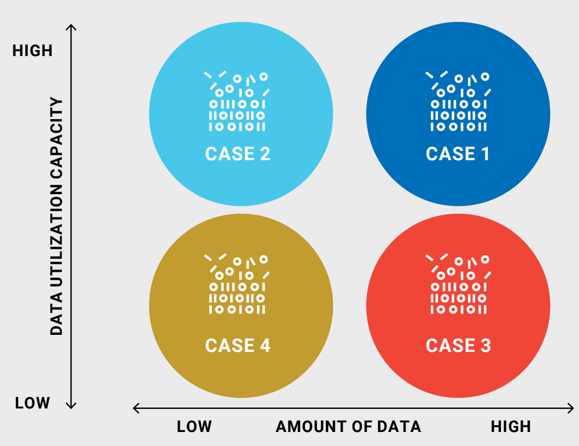
EVALUATION CATEGORIES AND EVALUATION RESULTS (3 LEVELS)	LEVEL 1 (WEAK)	LEVEL 2 (MODERATE)	LEVEL 3 (EXCELLENT)
Understanding of data standardization and technologies related to quality management			
Understanding of influence and degree of reflection of data analysis results in policies			
Solution that allows all types of consumers to freely use supplied data			
Preparation of policy utilization alternatives of supplied data / design of review guidelines			
Possession of anonymous information utilization technologies			
Permissible range of sharing for the convenient access of data (ex: permitted for all citizens, research institutions, local governments, etc.)			

4.4 Solutions per case for the advancement of Big Data-based policy utilization

There are limits to providing customized solutions for each country in the Asia-Pacific. However, according to the self-evaluations of the previous section, every country should fall under Cases 1 to 4. Hence, each country should be able to set up appropriate policies after considering the alternatives suggested in each case. The graph's horizontal axis, 'Amount of Data,' refers to the amount of public and private data the country owns. The graph's vertical axis, 'Data Utilization Capacity,' refers to a country's capacity to use a variety of qualitative data (which can be used to establish policies) through planning, collection, processing, analysis, and visualization.

This section suggests the roles of the country's public sector and private sector as well as their simultaneous role for each case and demonstrates a handful of feasible solutions and relevant case studies.

FIGURE 12 Four cases categorization to advance the Big Data-based policy utilization



Solution for Case 1 countries

Case 1 countries have high data utilization capacity and possess a large amount of data. Unlike other Cases, Case 1 can be seen as having some progress with the establishment of a Big Data ecosystem model of the public and private sectors. Therefore, this study aims to suggest solutions such as (1) (public) the establishment of a data governance system, (2) (public) the improvement of Open Platform's analysis functions and visual representation, (3) (private) the expansion of data connectivity of private data (which is not easily accessible by the public), (4) (public-private) the collection, generation, and processing of demand-customized data, as well as the creation of a suitable environment for data utilization, etc. The following case studies that are related to the solutions recommended above are presented.

1 View-T (The Korea Transport Institute, Republic of Korea)

View-T is a traffic monitoring and data supply-analysis platform that is being used in many fields (traffic policy, operation, planning, etc.) and for business and research purposes of various private firms and institutions, such as the Republic of Korea's central government, local governments, universities, etc. View-T was developed for the purpose of providing a foundation that allows for a deeper analysis of transport phenomena by facilitating the utilization of transport data for all citizens.

This platform was developed based on online services and overcame many challenges, such as (1) the issue of insufficient data collection in the public sector, (2) the issue of missing travel route data of vehicles, (3) the inconvenient collection and utilization of user data, and more.

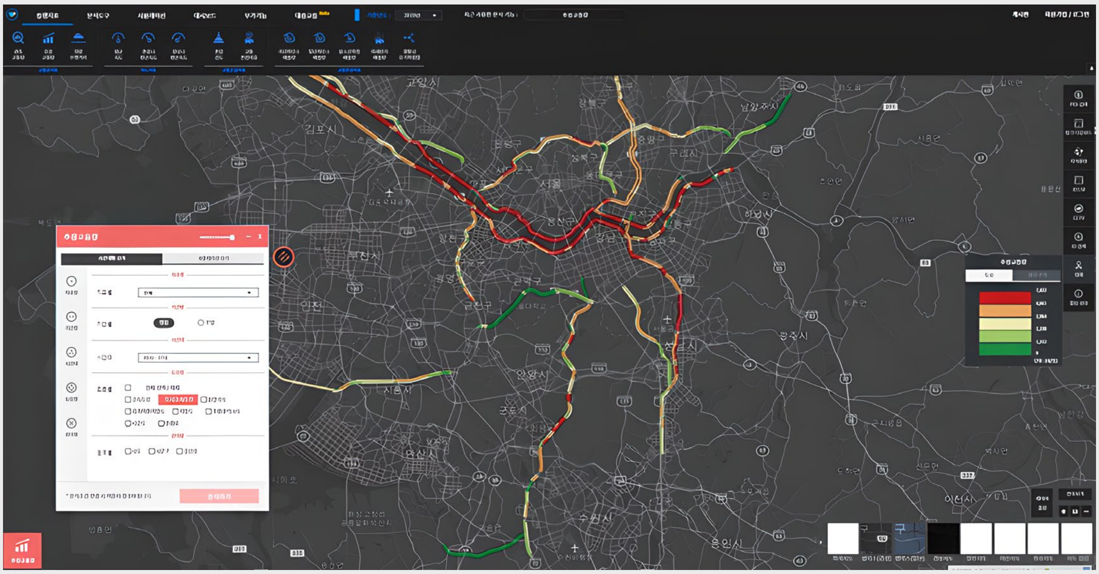
In order to resolve problem (1), View-T used private navigation data and public transport data to build traffic volume and speed databases of the entire country and secured over 95 per cent national coverage. To resolve problem (2), View-T actively connected private navigation data to develop a Big Data processing and fusion algorithm and allowed one to view the vehicle's actual travel route and mobility pattern. And to resolve problem (3), View-T proceeded to make developments that enabled data consumers to easily search for necessary information, to freely represent data on GIS bases, and to easily extract necessary data. This View-T development experience and functions (listed below) are anticipated to assist Case 1 countries' preparation of solutions for the advancement of Big Data-based policy utilization.

- Travel index: Uses mobility Big Data to produce structured calculations of road or spatial conditions, thereby providing objective and consistent status updates.
- Analysis tools: Provides data-based empirical analysis systems and environment through functions that allow temporal and spatial analyses of a mobility subject's (people, vehicles, etc.) travel characteristics and patterns.

Case 1 countries can develop platforms like View-T in the public sector to prepare a solution that enables all types of consumers to freely use and access data. Furthermore, the minimization of procedures between institutions and departments is anticipated for easier data sharing, as well as the improvement of data quality and utilization technologies best suited for the advancement of policy utilization through the connection between the public and private sectors.

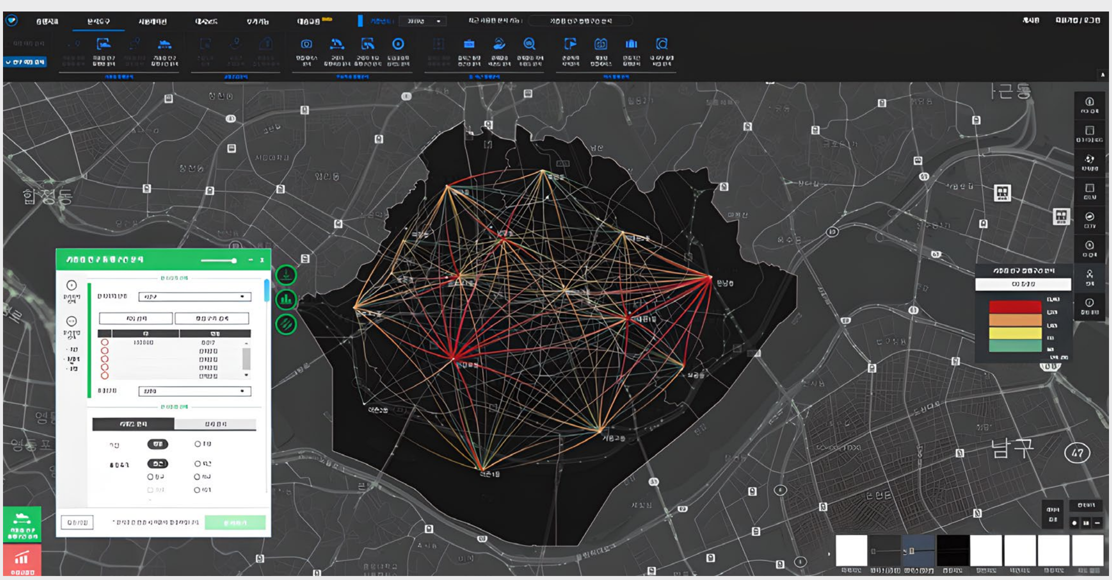
- Simulation analysis: Predicts time series-based travel pattern changes of the selected region per duration and time period through population, individual vehicle travel simulation analysis, etc.

FIGURE 13 Travel Index analysis of View-T



Source: Available at https://viewt.ktdb.go.kr/cong/map/second_map.do.

FIGURE 14 Analysis tool of View-T



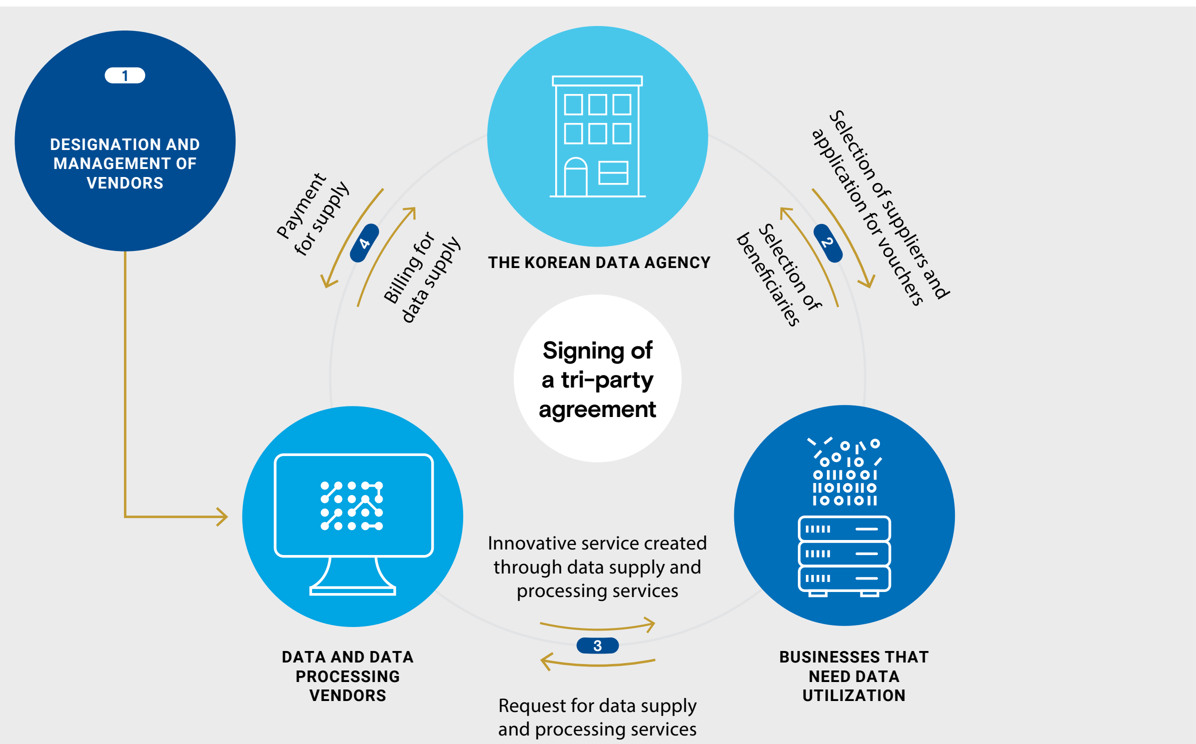
Source: Available at https://viewt.ktdb.go.kr/cong/map/second_map.do.

2 Data Voucher Project (Ministry of Science and ICT – Korea Data Agency, Republic of Korea)

The aim of the Data Voucher Business is to generate new products and services using data accumulated by companies and start-ups. This is a method where the government directly matches data processing companies and companies that possess basic data to produce new insights. As such, the role and strategy of the government are to support data processing throughout the entire industry, including cultural, transport, spatial, financial, and environmental industries, and to generate new jobs in the IT field.

The governments of Case 1 countries are suggested to commence data voucher businesses to create an environment that allows for the collection, processing, and utilization of data which corresponds to the country's data demand. This plan is beneficial for businesses that had ideas for data utilization but lacked funding as well as businesses without proper settings to sell data. As public institutions and departments carry out the role of a hub for transport-related data, the private sector's data trading framework and compensation system will become more defined. Furthermore, policy guidelines for data-based decision-making will be suggested at a more reasonable level.

FIGURE 15 Signing of a Tri-party agreement in data voucher business



Solution for Case 2 countries

Case 2 countries possess a low amount of data but have high data utilization capacity. This study reviews the problems of Case 2 by dividing them into two scenarios. The first scenario is when there is insufficient public data and a lack of proper data integration and connection with the private sector. The second scenario is when the capacity of the country's data collection system itself is lacking.

Thus, for the first scenario, this study suggests solutions such as (1) (public) the preparation of a policy for the integrated connection of data with the private sector, (2) (public-private) the preparation of a solution that enables the active exchange of data, technology, and labor with private firms specializing in data. For the second scenario, this study suggests (1) (public) the strengthening of potential through international and technological cooperation with other developed countries, and (2) (public-private) the planning of a nationally initiated project that enables the sustainable production of data even with a small amount of data by increasing the participation of the private sector.

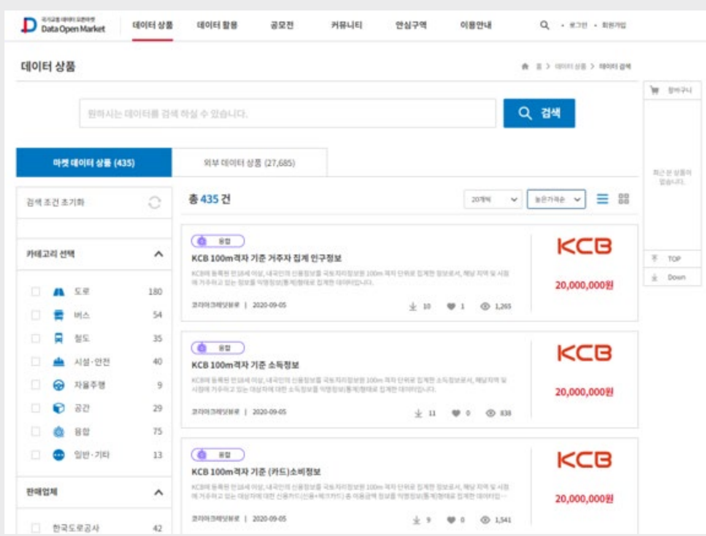
However, for these solutions to succeed, there should be no cases where data isn't shared due to the egoism between departments of public institutions, nor the monopolization or privatization of data due to the private sector refusing to cooperate with the public sector.

The following case studies that are related to the aforementioned solutions are presented below.

- National Transportation Data Open Market (Korea Expressway Corporation, Republic of Korea)

The National Transportation Data Open Market is a data transaction platform made through the cooperation of various road, railway, and public transport-related public institutions, local governments, and private firms that collect high-quality mobility Big Data. This platform (1) covers blind spots of data by integrating public and private data, (2) generates new data based on new technologies, and (3) aims to discover and expand new data-based businesses. It is important because it encourages the private sector to connect and share their data with the public sector by mediating the transactions of file format data between data suppliers and consumers.

FIGURE 16 National Transportation Data Open Market



Case 2 countries are recommended to increase the country's amount of possessed data by connecting the public and private sectors. Their respective governments should also actively promote businesses like the Data Open Market to create a successful data utilization environment. Furthermore, as Case 2 countries have high data utilization capacities of existing data in the private sector, countries should start R&D businesses that are capable of carrying out the examples in high-quality data utilization case studies, so that the public and private sectors can actively take advantage of these capacities.

Solutions for Case 3 countries

Case 3 countries possess both a low amount of data and a low data utilization capacity. Most Case 3 countries are expected to be developing countries in the Asia-pacific. Therefore, in this study, the following solutions are suggested: (1) (public) strengthening data collection capacities and expanding manpower, and (2) (private) strengthening data collection and processing technology capacities.

First of all, it is important to compare and analyze Big Data-related policies of other developed countries to derive implications and establish an environment in which advanced technologies can be supported. Through this process, policy alternatives can be prepared through feasible and suitable policy recommendations tailored to the characteristics of the recipient country. Ultimately, the way for fostering Big Data cooperation between countries can be paved, especially in the Asia-Pacific region, for humanitarian purposes.

Therefore, the following case studies are recommended as cases relevant to the solutions above. In particular, as can be seen in two cases, Big Data in the transport sector has the advantage of being able to grasp the 'mobility' of the entire region and nation continuously and accurately. Through this, the data can be used to establish various customized policies related to the improvement of urban structure based on mobility characteristics. This includes the travel behavior analysis of cities, the identification of regions and socioeconomic groups that lack modes of transport, and balanced regional development. It also becomes possible to create new policies and enhance many data-based transport operation technologies, in order to increase the efficiency of signal operations, congestion management, and public transport systems.

1 Ukraine Smart City project: Data Utilization Technology Development Assistance Case[6]

Despite having data that could be collected through CCTV and small amounts of urban spatial data, the municipal government of Kyiv, Ukraine, had trouble deriving traffic engineering insights. In order to overcome these limitations, the Seoul Urban Solution Agency proposed to apply the shortest route algorithm and trajectory analysis technology as a part of Kyiv's Smart City project in 2018. A technology that could generate a vehicle's traveled route and augment information with CCTV data and a technology that could monitor road congestion were developed.

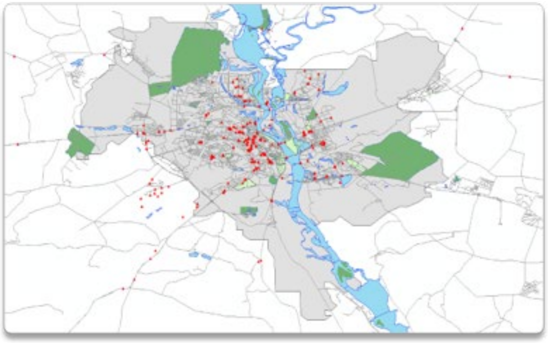
In the past, CCTVs were only installed on some roads, so even if the vehicle passed through each CCTV continuously, the road and time the vehicle passed through couldn't be identified. To solve this problem, the route that was most likely used was estimated, followed by the calculation of the vehicle's entire route, and based on this data, a database was built. As a result, the cases were classified into three types according to the pattern of traffic volume passed through each CCTV during three different time periods.

This is significant in itself because it overcame the limitations of CCTV data, which could only indicate whether a vehicle had passed a certain point or not. There are two ultimate implications: (1) while developing an estimation algorithm, the research team discovered that it was possible to monitor the entire city's travel speed even with a small amount of data, and (2) with this small amount of data from Kyiv City and the technological capabilities of advanced countries, it was possible to suggest ways to promote policy utilization at the national diplomacy level.

6 Seoul Urban Solution Agency, 2018.

FIGURE 17 Overview of CCTV data

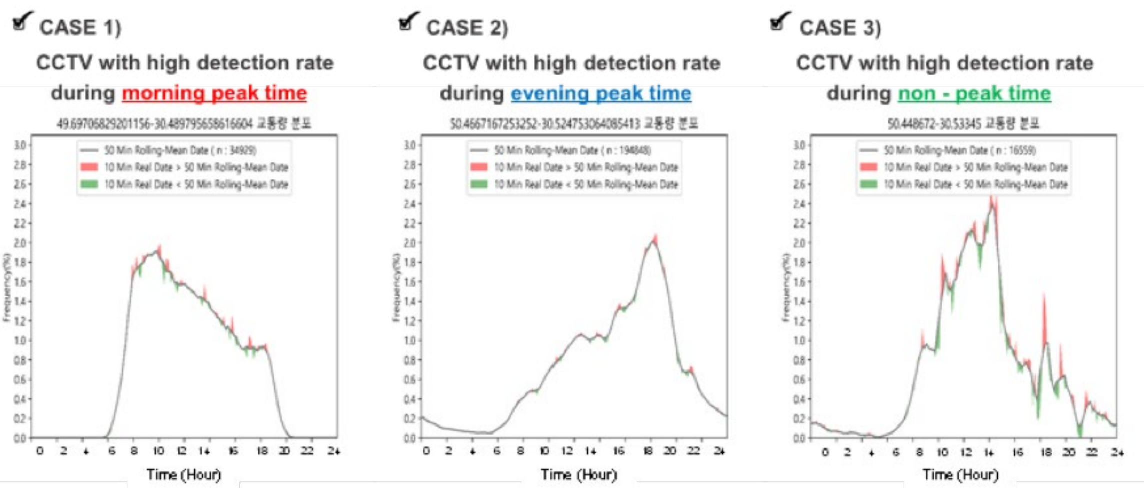
KYIV CCTV INSTALLATION STATUS



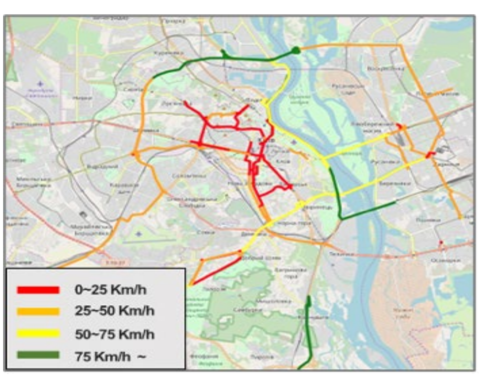
INFORMATION COLLECTED THROUGH CCTV (VEHICLE ID, TIME, LOCATION)

num	id	time	lat	lon
13944137	AA1682AC	437573/50.43770332420319/30.518399885442047		
13944138	CA7448CE	437573/50.4410567/30.5496817		
13944139	GD0069T	437574/50.46675770042869/30.52468869106906		
13944140	AA9731PX	437573/50.44056569500019/30.627996603379383		
13944141	BHM272	437574/50.44340378617667/30.43943381904		
13944142	AI3094ET	437573/50.44944129258708/30.47718942657049		
13944143	AA4041M(	437574/50.45124694099228/30.52289429323896		
13944144	AI3094ET	437573/50.44948228265824/30.47719479098867		
13944145	AA5910PN	437573/50.440637441055955/30.62792686594485		
13944146	AI1816EK	437573/50.38123268322768/30.47644376284422		
13944147	AA2674PC	437573/50.44948228265824/30.47719479098867		
13944148	AC5578BC	437571/50.441474/30.16479		
13944149	AA9695MI	437573/50.43703946728385/30.49161851422062		

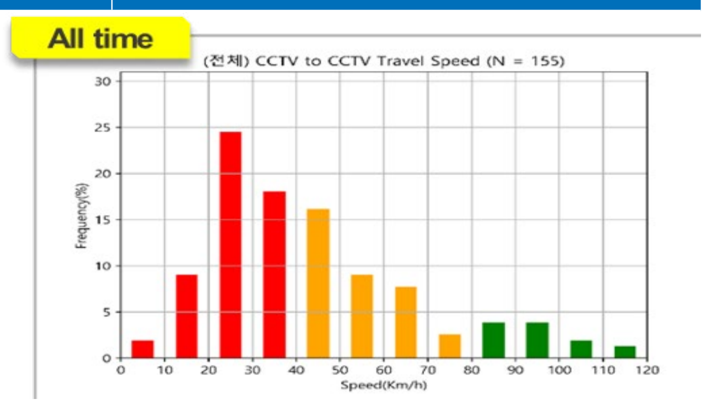
FIGURE 18 Analysis results of three cases



ESTIMATION OF THE TRAVEL SPEED BY ROAD SECTION IN KYIV



FREQUENCY BY ROAD SECTION SPEED (%)



2 Mexico City KSP (knowledge sharing program): Data Construction for Improving Public Transportation Networks[7]

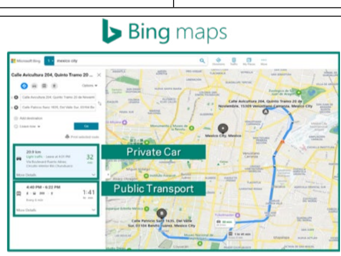
KSP is a knowledge-based development project that provides customized policy suggestions to partner countries based on the Republic of Korea's experience and knowledge of economic development. Its objective was to come up with a measure to enhance Mexico City's public transport network. Through this business the researchers proposed data collection and analysis technologies to Mexico City's Transportation Department (SEMOVI, Secretaria de Mobilidad), which would allow for the investigation of areas that lacked public transport based on existing transport-related data.

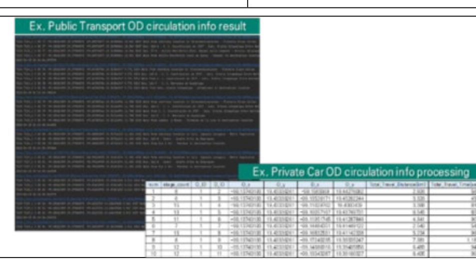
EOD Data (OD circulation survey data) was the only data that could be used to analyze citizens' travel characteristics in this research. Through this data analysis, information such as 1) estimation and comparison of weekday and weekend traffic, 2) identification of transport modes used to reach the final destination, 3) identification of peak traffic, and 4) purpose of travel, time, and cost could be derived. However, there were difficulties analyzing areas lacking public transport.

An analysis technology from Bing Maps, which collects and combines travel information per mode, was proposed. This technology uses Web Crawling, which automatically investigates data that exists on the web, and directly collects and stores data according to the user's preferences.

FIGURE 19 Web-crawling methods by using map API (e.g. BingMaps)

DATA NAME	TIME SPAN	SPATIAL EXTENT	CONTENT SCOPE
EOD	23 January 2017 – 3 March 2017	ZMVM(Metropolitan Zone of the Valley of Mexico)	Identification of traffic characteristics of 66,625 households (purpose, time, modes, cost, etc.)
Privately-collected traffic information (BingMaps)	13 April 2022 – 14 April 2022 (collection period)		Identification of actual traffic flow characteristics for 7,150 OD pairs in ZMVM (time, modes, route, etc.)





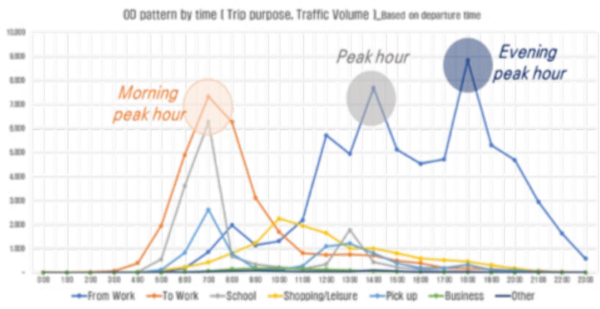
CATEGORY	CONTENTS
Time Span	13 April 2022 – 14 April 2022, available at any time
Spatial Extent	7,140 OD pairs in CDMX in the Metropolitan Zone of the Valley of Mexico (ZMVM)
Purpose	Acquisition of detailed traffic information such as actual travel time, modes, and route
Necessity	Comparison and analysis with traffic characteristics in 2017 (before COVID-19)
Subject of Investigation	Real-time traffic flow of 7,140 OD pairs via BingMaps API
Investigation Details	Acquisition of real-time traffic information (using API)

It was discovered that through EOD data, analysis results from a macroscopic perspective could be derived, and that (1) OD pattern by time and (2) Modes-sharing ratio could be understood.

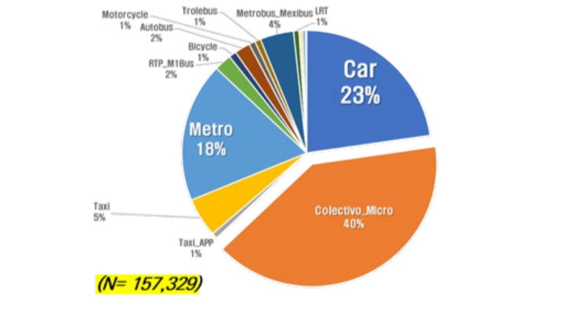
7 Studio Galilei Co., Ltd., 2022.

FIGURE 20 Basic statistical analysis results

(1) OD PATTERN BY TIME



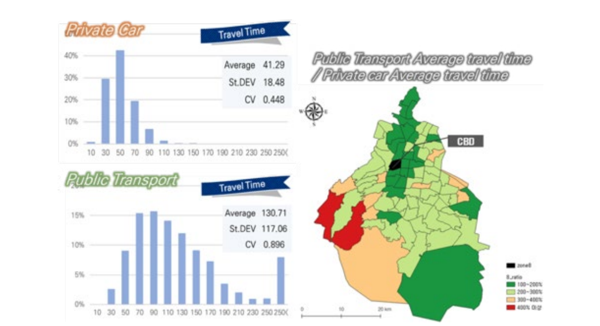
(2) MODES-SHARING RATIO



In addition, the data collected from Bing Maps was able to derive both analysis results of a macroscopic perspective and analysis results of a microscopic perspective. Therefore, (1) Comparison of passengers car-public transport access level, (2) Movement confidence analysis, and (3) Public transport vulnerability (Zone, OD based) that couldn't be analyzed in the past were learned. Through this, the researchers were able to provide insights regarding the direction policies should take to assist areas lacking public transport.

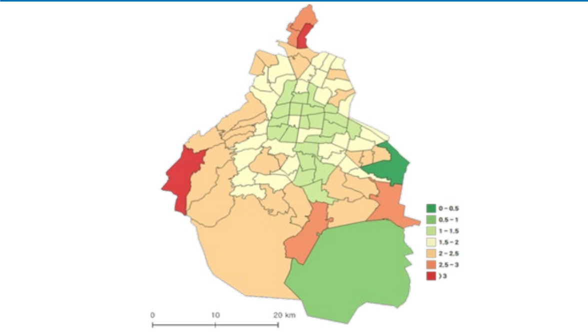
FIGURE 21 The analysis results for the level of mobility service level

(1) PASSENGER CAR-PUBLIC TRANSPORT ACCESS LEVEL



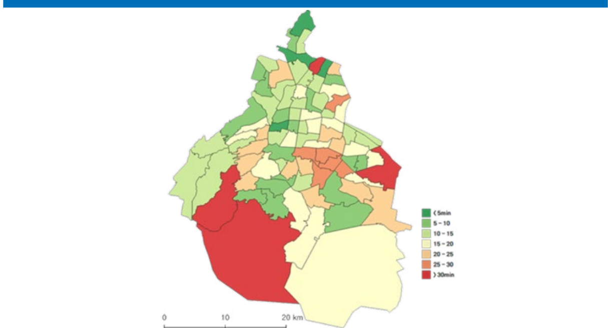
Public transport mobility equity and spatial equity are relatively weak (280% of travel time related to private car's)

(2) ANALYSIS OF MOBILITY CONVENIENCE (STAGE*, NUMBER OF TRANSFERS)



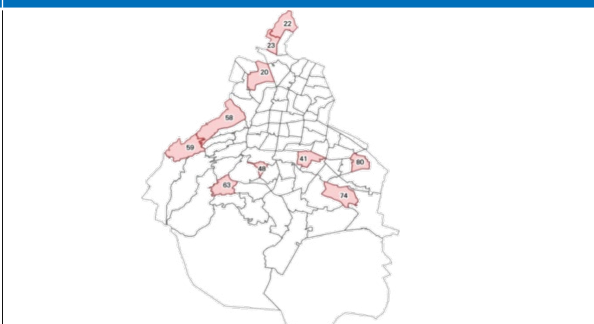
The average number of transfer is 1.89 in Mexico city when excluding 'only walking(stage=1)'

(3) MOVEMENT CONVENIENCE ANALYSIS (FIRST MILE TRAVEL RATIO)



The average total walking time takes around 30.34 % of entire stage in Mexico city

(4) VULNERABLE ZONES USING PUBLIC TRANSPORT



These 10 vulnerable zones are urgent to reduce the number of transfer and stage

Note: *=In 2015, when the concept of MaaS was not widely known, the Stage indicator presented in the MaaS Feasibility Assessment in London, UK proves that it is a very effective indicator for evaluating the level of seamless transit service for the entire city.

This is significant because it supplemented the limitations of the low level of traffic-related data in the existing Mexico City. There are ultimately two implications regarding this point: (1) the research team not only suggested a technology that could collect private sector travel information data, but also identified the citizens' quantitative mobility level with a small amount of data and 2) was also able to suggest a feasible solution to increase the acceptance for data-driven policies based on the technologies of developed countries and Mexico's small amount of data.

Solution for Case 4 countries

Case 4 countries possess a high amount of data but have a low data utilization capacity. Case 4 countries, just like Case 3 countries, are expected to be developing countries in Asia Pacific. In this study, the following solutions have been suggested for these countries: (1) (public) strengthening data utilization capacities and expanding data accessibility, (2) (public) active investment of private companies in the data industry, and (3) (public-private) promoting active interactions between technology and human labor.

First of all, it is necessary to expand the country's technological capacities and improve the intrinsic value of data through active technological interactions with the private sector. In addition, the public sector should make data more accessible to provide an opportunity to actively derive insights from individuals, the private sector, and academia, like the case of AURIN. If a country's self-sustainable development is deemed to be difficult, it would be ideal to establish an environment where the country can receive support (e.g., technological cooperation, potential strengthening programs, etc.) from developed countries.

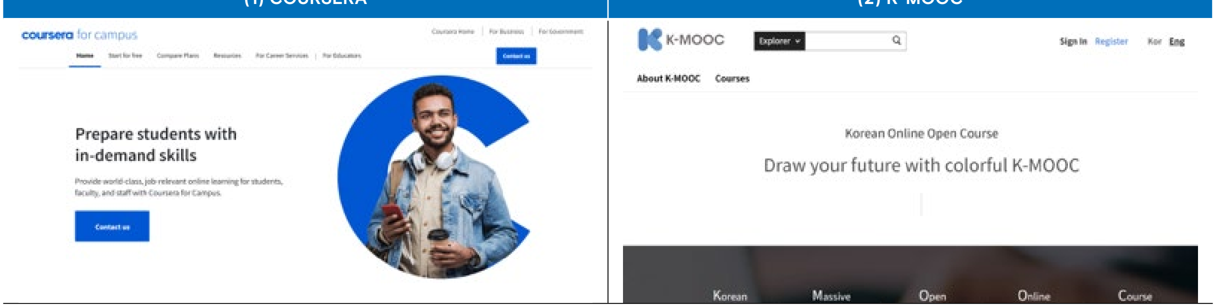
Moreover, from a mid- to long-term perspective, an effective solution for Case 4 countries may be to strengthen their training curriculum and data analysis systems. Although educational efforts to cultivate young intellectuals are important, expert education for current workers may be more effective for achieving results in a short period of time. Case studies for establishing and improving a data-related education system are recommended as follows.

1 Coursera

Coursera is a Massive Open Online Course (MOOC) platform founded in 2012 by Professor Andrew Ng and Professor Daphne Koller of Stanford University's Department of Computer Engineering. As lectures are provided free of charge, if the country's educational environment is poor, the platform can be utilized to form an educational curriculum related to data utilization. It is also possible to adopt similar platforms to suit each country's education system. In the case of the Republic of Korea, its own platform called K-MOOC commenced in 2015, and it is operated and managed with the support of the National Lifelong Education Center.

FIGURE 22 Training curriculum for data analysis friendly systems

(1) COURSERA	(2) K-MOOC

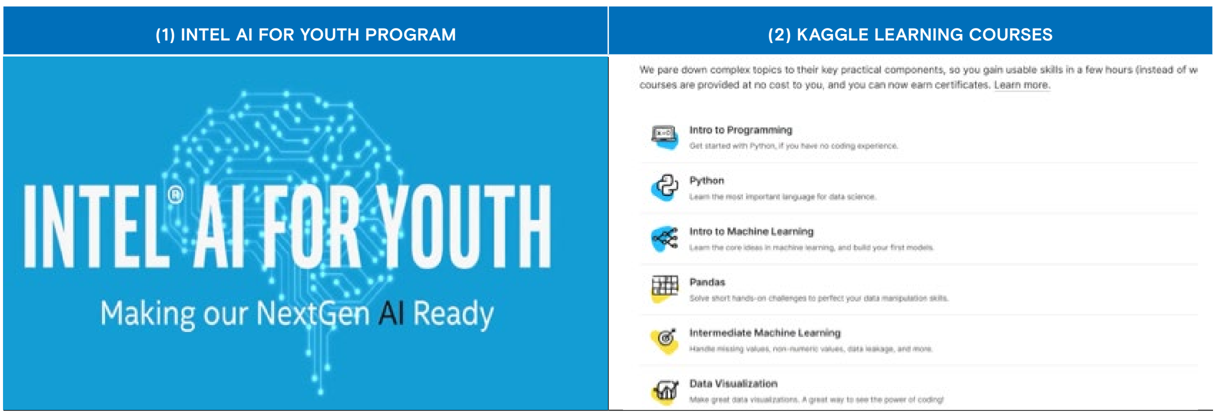


Source: Available at https://www.coursera.org/ (01 Sep 2023) & Available at http://www.kmooc.kr/eng (01 Sep 2023).

2 Data analytics and AI-specific training programs

Coursera is a platform that not only enables data analysis but also provides college-level education. The Intel AI for Youth Program introduced in this section, however, specializes in data analysis learning. This program helps elementary and middle school students, as well as those that have no experience with AI and data science to better understand these topics. As of now, is being used in 11 countries, including The Republic of Korea, India, and Poland. The introduction of programs like this are strongly recommended because they can be easily adopted once they are translated and reformatted for the respective country. In addition, famous data competition platforms such as Kaggle have recently opened various courses for data analysis. Since all courses and source code are free, if one uses the platform, highly efficient results can be achieved at a low cost.

FIGURE 23 Data analytics and AI-specific training programs



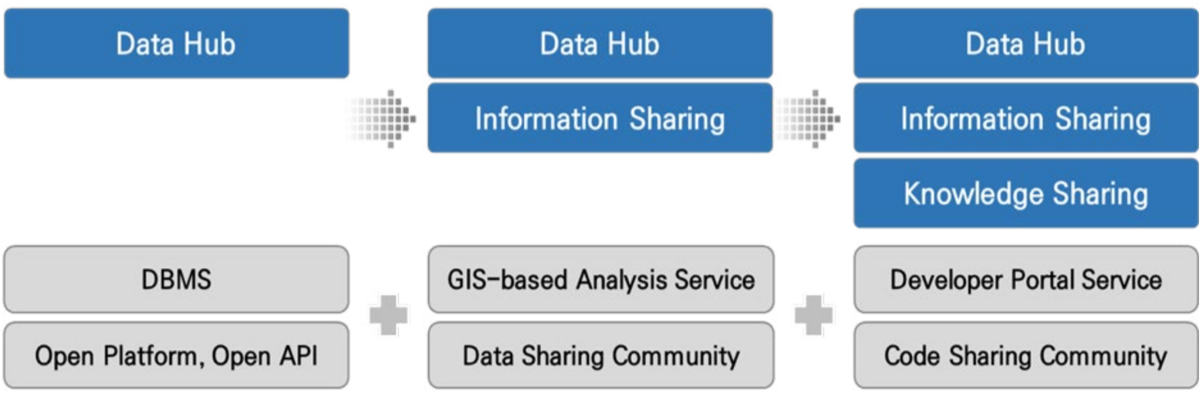
Chapter 5

Big data utilization goals and policy proposals of Asia-Pacific countries

5.1 Suggestion of Asia-Pacific countries' Big Data utilization goals

There are 3 steps to suggesting each country's Big Data utilization goals. Step 1 (short-term) is whether or not there is a 'Data Hub' (e.g., Open Data Portal) opened by the public or private sector in the country. Step 2 (mid-term) is the progress of 'information sharing' and whether it is on a level that allows for the connection and fusion of various heterogeneous data and active utilization by the country. Finally, Step 3 (long-term) is whether 'knowledge sharing' has been produced, specifically for the purpose of creating a widescale Big Data industry.

FIGURE 24 Big Data utilization goals with three stages



1 Step 1 (short-term): An environment that enables data sharing through the building of a Data Hub

Public data is categorically classified and managed through each country's Open Data Portal. Although each country's specific categories for transport data might differ due to the disparities in their level of data fusion as well as how much object/agent, socioeconomic, and spatial data they each possess, it is best that they manage the data through a publicly led Open Data Portal and be as transparent as possible with their data. Many countries that possess data are trying to activate and manage their Open Data Portal for this purpose. This produces benefits for consumers, such as citizens and professionals in various fields, by making their lives more convenient.

2 Step 2 (mid-term): An environment that enables information fusion and cooperation for Information Sharing

Rather than just owning data, insights should be drawn from data analysis to satisfy users' needs (whether public or private) and advance towards the goal of knowledge sharing. Data analysis is particularly helpful for deriving diverse and effective insights through data fusion. Additionally, data usability can be greatly increased, as data is crucial for comparing solutions and making final decisions. Countries that have reached this stage through open data initiatives have successfully created an environment that not only provides information but also allows for better decision-making, including research partnerships with academia.

3 Step 3 (long-term): An environment with a widescale Big Data industry through knowledge sharing

The last step of Big Data infrastructure building is the creation of a Big Data industry. This refers to the establishment of a data environment where the building of public Big Data not only involves Big Data generated from the public sector but also allows anyone to freely use and access data purchased from the private sector to draw new insights for public interests and enhance data quality. For example, the Republic of Korea has developed and is running a transport/Geo-data open market, with a currently active road traffic data market. Singapore is also actively exchanging data with the private market through various routes. Other developed countries are working to create a digital economic environment as well.

Countries such as the Republic of Korea, and Singapore, where their level of data integration is considered 'mature,' correspond to Step 3, while the majority of other Asia-pacific countries are at Step 2.

5.2 Policy proposals for the advancement of Big Data utilization

This study proposes countries or institutions to strengthen their Big Data utilization potential through four steps. (1) Step 1 is to identify the country/institution's necessary data potential, (2) Step 2 is to evaluate the data potential of the country/institution's members, (3) Step 3 is to analyze the country/institution's disparities in data potential, and (4) Step 4 is to begin preparing a solution to strengthen data potential.

Step 1 considers the corresponding country/institution's vision, goals, policy priorities, etc., to define the country/institution's necessary capacity to utilize data and the required standards per key performance indicators (of the capacity) in order to achieve the organization's goals. Step 2 measures the current capacity of the country/institution's members to use data according to what was determined to need strengthening in Step 1. Step 3 analyzes the difference between the data-utilization capacity the country/institution could aim for and its actual capacity to use data in the present. Finally, Step 4 is where a solution is established and implemented to resolve this disparity.

Each country is encouraged to work towards the advancement of Big Data utilization through these four steps: identification, evaluation, analysis, and solution preparation. In the previous chapter, the contents of steps 1 to 3 were reviewed, and thereby proposed case studies and appropriate solutions for each case. In this section, situation-based solutions are proposed to reduce the disparities in the country/institution's data capacity. Lastly, if results of steps 1 to 4 are gathered from Asia-Pacific countries through common procedures and forms, better policies may be proposed in the future.

TABLE 20 Four steps for the advancement of Big Data utilization

STEP 1	STEP 2	STEP 3	STEP 4
Measure the data-utilization capacity the country/ institution should aim for	Evaluation of the country/ institution's members' capacity to use data	Disparity analysis of country/ institution's capacity to use data	Preparation of a solution to strengthen the country/ institution's capacity to use data
1 Select capacity while taking into account the country/institution's vision, goals, and policy priorities 2 Define key performance indicators for building the capacity to use data 3 Write a definition of the country/institution's necessary capacity to use data	1 Distribute a survey with all members of the country/institution as the sample population 2 Write an evaluation of the current data utilization potential	1 Conduct a comparative analysis between the evaluations of the necessary data-utilizing capacity and current data-utilizing capacity to calculate the capacity the country/ institution needs 2 Rank each key performance indicator by priority while considering the country/ institution's goals and chances of implementation	1 Devise a solution to reduce the disparity in the country/ institution's capacity to use data 2 Promote the strengthening of the country/institution's capacity to use data
RESULTS			
Definition of necessary capacity to use data	Evaluation of current capacity to use data	Rank KPIs by priority to strengthen the country/ institution's capacity to use data	Solutions to strengthen the country/institution's capacity to use data per KPI

When there are inevitable utilization limitations due to differences in development between countries

Depending on the country's level of development, there can be huge differences in the status of public data building and utilization cases of such data. Thus, whether each country has an environment that enables Big Data to be utilized in policies should be reviewed.

- The public sector could achieve the practical and reasonable utilization of Big Data in a sustainable manner.
- The country itself needs to review Big Data case studies (experiences) of other countries and the obstacles they faced.
- In addition, although many countries are already generating data, there might be countries without an Open data portal due to the lack of manpower and utilization systems.

Limitations due to the absence of an integrated data management system and organizational cooperation within the country

As there are cases where collected data is separated and managed by multiple subjects, there needs to be a strategy that enables experts to manage integrated data. Though there might be an environment ready for policy use, a detailed diagnosis could be conducted on whether it is being managed efficiently by the country.

- After preparing a clear national mid-long-term vision, the rationalization of data obtainment procedures, strengthening of data accessibility, enhancement of openness measures (reformation of legal policies), etc., are essential.
- Active cooperation to prevent data privatization and monopolization between data collection, operation, and management institutions and related departments is necessary.
- Evaluation and diagnosis of public data quality, support of enhancement efforts, standardization, education, training, etc. are also necessary.

Limitations of public and private data integration or the enhancement of data's intrinsic value and quality

The development of an innovative strategy that reflects data-driven decision-making and is also simultaneously successful would be challenging to achieve with publicly managed data alone. Therefore, there is a clear need to search for a solution that produces practical values fitting for the actual utilization purposes of public data.

- (Quantitative) Need to deduce a more expanded Insight by linking various fields amongst publicly managed data and putting them into consideration.
- (Qualitative) Enhancement of data quality through opening public data and utilization support.
- (Optimization of Personalized Data for Better Usability) Need to search for ways to obtain individualized data that match the actual needs and preferences of citizens to supplement public data, as usability could decrease with only statistical data.
- In other words, countries need to enhance the intrinsic value of the Big Data they possess by establishing a form of data supply that considers user demand rather than a supplier-centered form of data supply.

Chapter 6

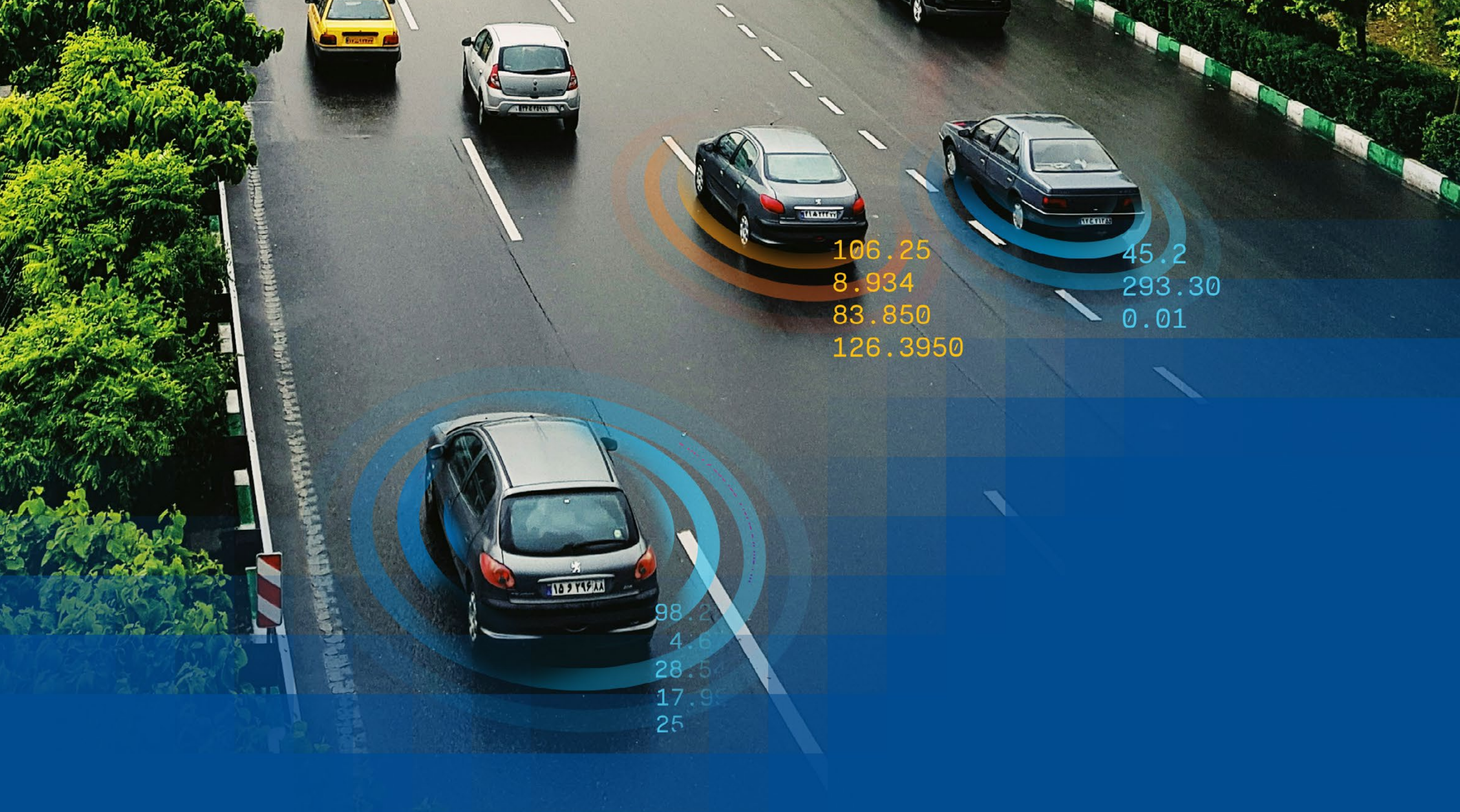
Recommendations

Sustainable Development Goals (SDGs) explain the roles and responsibilities of developed countries in achieving fair and equal development in all dimensions and allow developing and developed countries to have a shared sense of responsibility. As a result, interest for data-based socioeconomic development has increased throughout the world. This gradually raised the necessity of generating consistent and effective data at a more specific level. However, in reality, many developing countries often do not collect data for analysis even at the most basic level. Besides, even if they do, a substantial number of countries still lack the capacity to use this data properly. Therefore, such countries need to solve these issues regarding their data capabilities. Hence, this study examined the limitations many Asia-Pacific countries have experienced or are experiencing regarding the collection, processing, and utilization of Big Data. This report's ultimate goal is to suggest a guideline and direction for countries as they promote Big Data utilization and advance to Big Data-based policy decision-making. In the end, what is required of countries are collection systems and skills for the creation of a Big Data-based environment. There should also be a reformation of governance systems, so that various subjects (e.g., the government, private firms, non-government institutions, etc.) are able to construct networks. This would need to accompany international and intranational efforts, such as the following suggestions below.

Firstly, public Big Data and private Big Data could sustain an ongoing interaction. Thus, the creation of an environment where various Big Data owned by the private and public sectors can be integrated, easily accessed, and shared for users is extremely important. More specifically, countries would have to develop a universal and easily accessible data structure available to all users, while also applying a structure to improve data credibility. It would also be more desirable to create a setting where citizens can use this data for free, rather than allowing private firms to earn a lot of profit through transport Big Data. To achieve this, the government needs to establish a Big Data governance system, and each country would have to secure a budget for the sustainable operation of a data industry and devise a plan to create an organization responsible for the task.

Secondly, countries could develop a variety of transport operation technologies to minimize the transport sector's socioeconomic costs by using an integrated Big Data built by the public and private sectors. If the country is incapable of accomplishing this alone, they should form global partnerships through international and technological cooperation with countries that own analysis technologies. Through capacity strengthening training, policy consultation, and sufficient support for building a Big Data center, developing countries would be able to strengthen their Big Data utilization capacities. In order for this to happen, each country is suggested to begin with the most basic self-evaluations, such as understanding their data collection laws, data sharing and utilization limits, and the level of data interaction between their public and private sectors.

Ultimately, to be able to actively use Big Data in policies, a country cannot simply possess Big Data's technological capabilities. Instead, the country could consider national strategic concerns, such as improving their ability to draw insights through Big Data. The key to this process is the cultivation of experts who are able to analyze and manage Big Data. However, this is an issue that cannot be solved by the government alone and is a problem that has to be overcome through partnerships with private firms or experts within or outside of the country. Meanwhile, in a broader sense, these abilities entail not only the learning of data processing abilities on an individual level, but the equipment of Big Data intelligence on a national level. When viewing the problem from this perspective, countries could consider policies that involve the installation of a national Big Data center or the employment of professionals who can manage Big Data.


106.25
8.934
83.850
126.3950
45.2
293.30
0.01