

UTILIZATION OF NON TRADITIONAL DATA

Dr. Edi Setiawan, MSE, MSc

Head of Population and Labour Mobility Statistics Division
BPS-Statistics Indonesia



28 SEPTEMBER 2020



Outline

- Introduction of Big Data
- Mobile Positioning Data (MPD) and BPS' Initiatives with MPD
- The Use of Satellite Images and Android Based Application with Digitalization Data Processing for Area Sampling Frame
- The Use of Google Mobility





BIG DATA



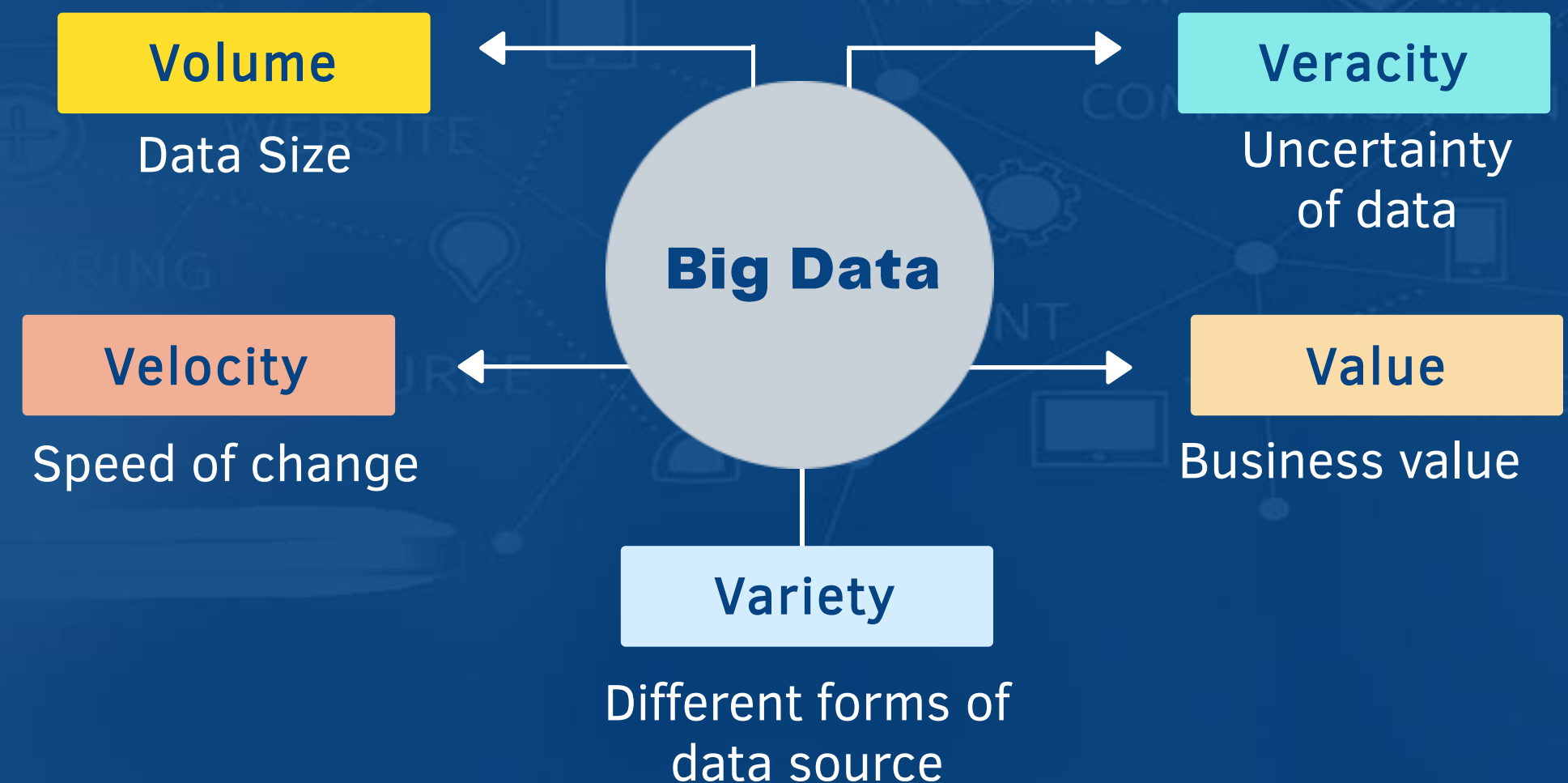


“What is Big Data?”

Big data is non traditional data that characterized by **large volumes of different types of data** (e.g. Social, web, transaction, etc) That build very quickly.

It exceeds the reach of commonly used hardware environments and software tools to capture, manages, and process in timely manner for its users.

”



How Big is Big Data

- **2,9 Million** number of emails sent every second
- **375 Megabytes** data consumed by households each day
- **20 Hours** video upload to Youtube every minute
- **24 Petabytes** data per day processed by Google
- **50 Million** tweets per day
- **700 Billion** total minutes spent on Facebook each month
- **1,3 Exabytes** data sent and received by mobile internet users
- **72,9 Items Products** orders on amazon per second

Media

Media and communication outlets (articles, podcast, audio, video, email, blogs)

Social

Digital material created by social media (text, photos, videos, tweets)

Machines

Data generated by computers and machines (business process log, sensors, phone calls)

Historical

Data about our environment (weather, traffic) and archived documents, forms or record





Opportunities on using



A huge potential for various statistics production

People using technology in daily life

There are several legal aspect for utilizing big data

To reducing cost/budget and respondent burden

To producing timely and real time information

To improve data sampling frame and estimates from statistical sources

To provide complementary statistical information



MOBILE POSITIONING DATA (MPD)



Facts of **MPD in Indonesia**



355 million active telephone numbers vs 268 million population (2019)



All numbers have to register based on Single Identification Number as in the personal ID Card



91% of the adult population using mobile phone device (any type) and 60% using smart phone (2019)



Benefit of Using MPD



INCREASE COVERAGE

Mobile phone data is used to increase the coverage data.

LESS BURDEN

Less work and respondent burden

COST EFFECTIVE

Less budget than traditional survey
(1 vs 8 million US\$)

GRANULAR DATA

More granular data, such as municipality and sub-district

TIMELY

More timely and up to date data than traditional survey

MORE FREQUENT DATA

Annual data for monitoring



Mobile Network Operator (MNO)



- **Call Detail Record (CDR):** mobile-call, mobile-sms
- **Location Based Service (LBS):** Internet Protocol Detail Record (IPDR), Location Area Update, Handover event, device turn on, and many more



MPD Data Source

subscriber_ID	start_date	end_date	kecamatan_name	kabupaten_name	province_name
(1)	(2)	(3)	(4)	(5)	(6)
628111*****	12/1/2018 9:37	12/1/2018 14:28	PANCORAN	KOTA JAKARTA SELATAN	DKI JAKARTA
628111*****	12/1/2018 15:09	12/1/2018 16:29	SETIA BUDI	KOTA JAKARTA SELATAN	DKI JAKARTA
628111*****	12/1/2018 18:08	12/1/2018 23:59	TANAH ABANG	KOTA JAKARTA PUSAT	DKI JAKARTA
628111*****	12/2/2018 4:40	12/2/2018 4:59	CIBODAS	KOTA TANGERANG	BANTEN

WHO

WHEN

WHERE

Travel duration

travel distance between regions

modes of transportation



MPD Data Structure



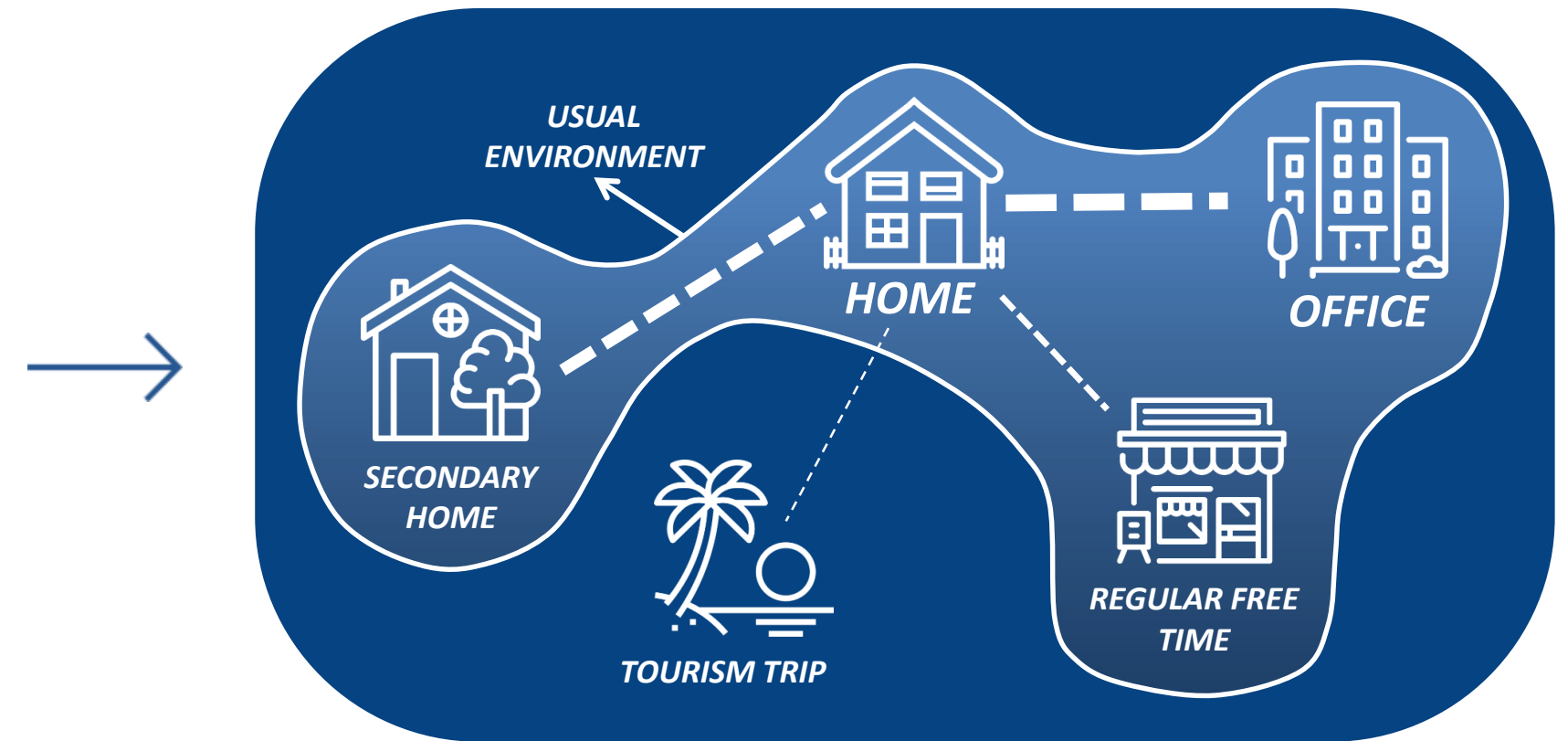
Mobile Positioning Data (MPD)

01

Usual Environment

- **Home:** phone number that is most exposed in a location at night (5.00 PM until 8.59 AM at next day)
- **Work:** phone number that is most exposed in a location from morning to evening (9.00 AM to 4.59 PM on weekdays)

others: secondary home, regular free time, tourism trip



02

Method: Sandboxing

- Access database by using a token from the MNO
- Limited only to the process of data exploration and algorithm development.
- Limited access to sample data, not overall data.
- MNO built a script to be implemented on the data.
- Running the script

03

Quality Assurance

- Validation for trip identification results
- Validation using volunteer data
- Validation result used to improve methodology

01



BPS' Initiatives with MPD For Official Statistics



BPS' Initiatives

with MPD



Foreign Tourist at Cross Border Posts



Domestic Visitor



Event based analysis:

- Asian Games
- IMF-WB Annual Meeting



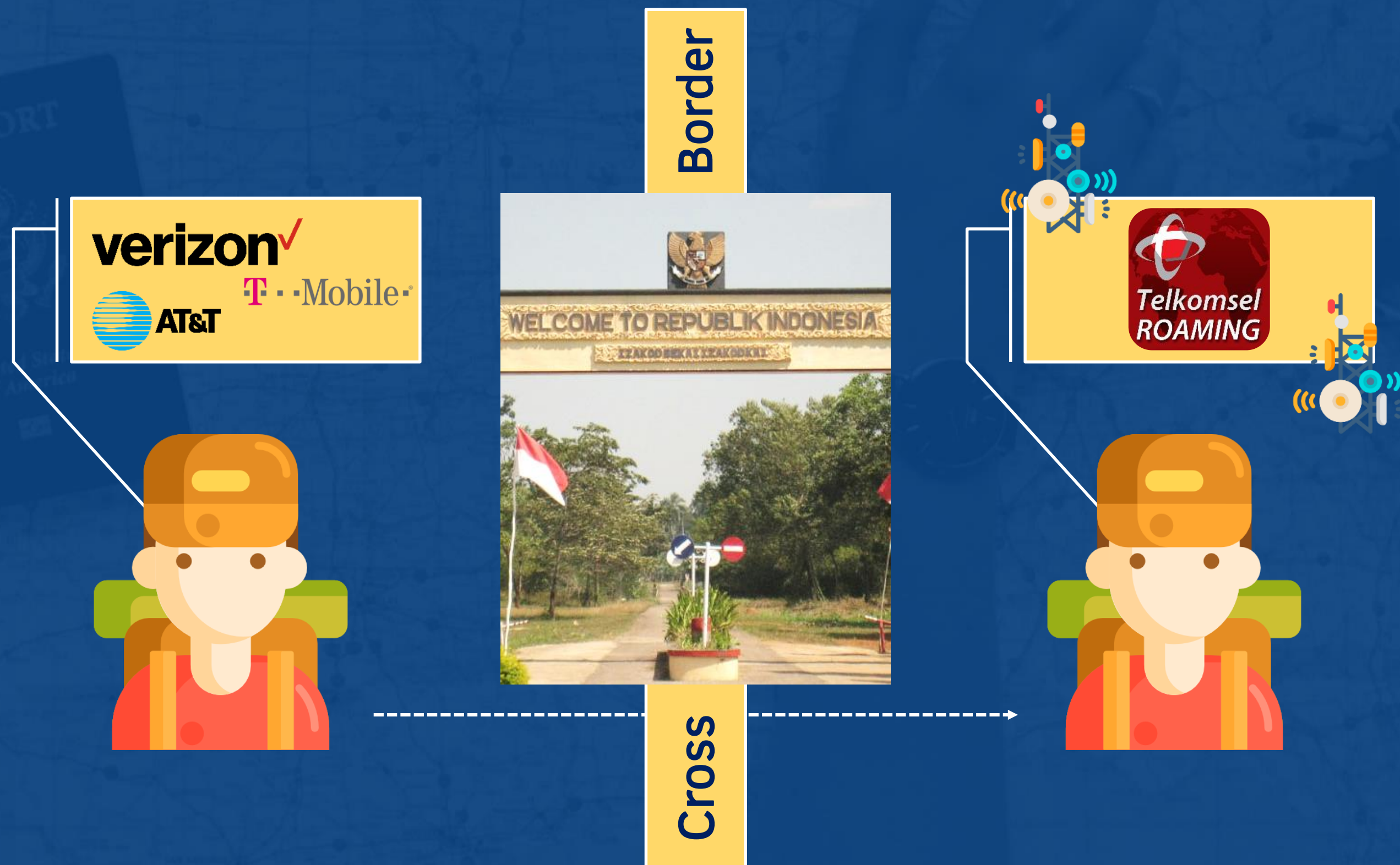
Commuting Statistics



Metropolitan Statistical Area



Foreign Tourists at Cross Border Posts

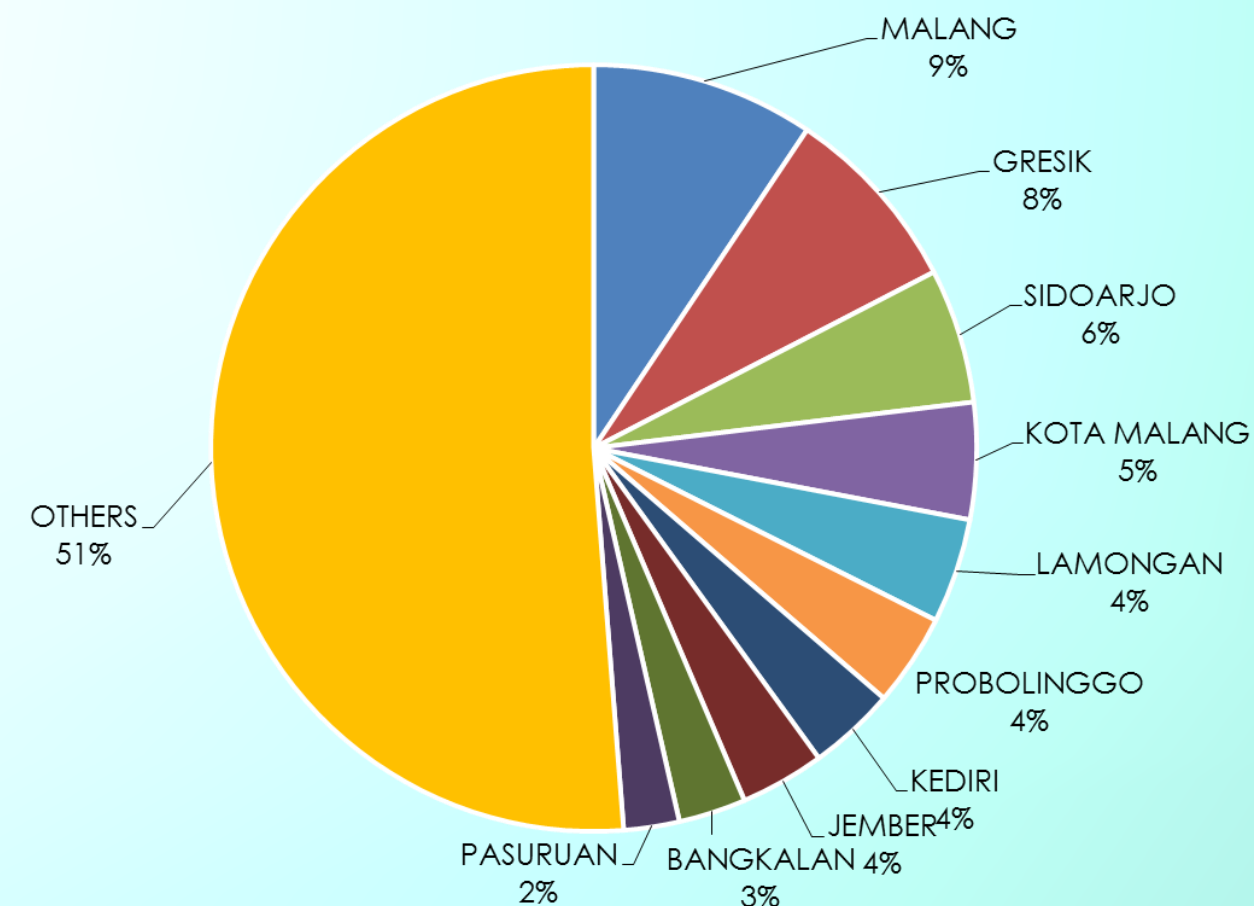




Domestic Tourism

- MPD using to correct survey data
- Coverage: 190 million subscribers, all Indonesia
- Time Reference: Monthly
- Output: O-D Matrix of 514 district/city
- Expenditure data: digital survey 50.000 travellers

Domestic Tourist to Surabaya





ASIAN GAMES | 2018
Jakarta Palembang



ANNUAL MEETINGS
2018 | indonesia
INTERNATIONAL MONETARY FUND
WORLD BANK GROUP

Event Based Analysis

MPD vs Commuter Survey

on Jabodetabek Area

Criteria:

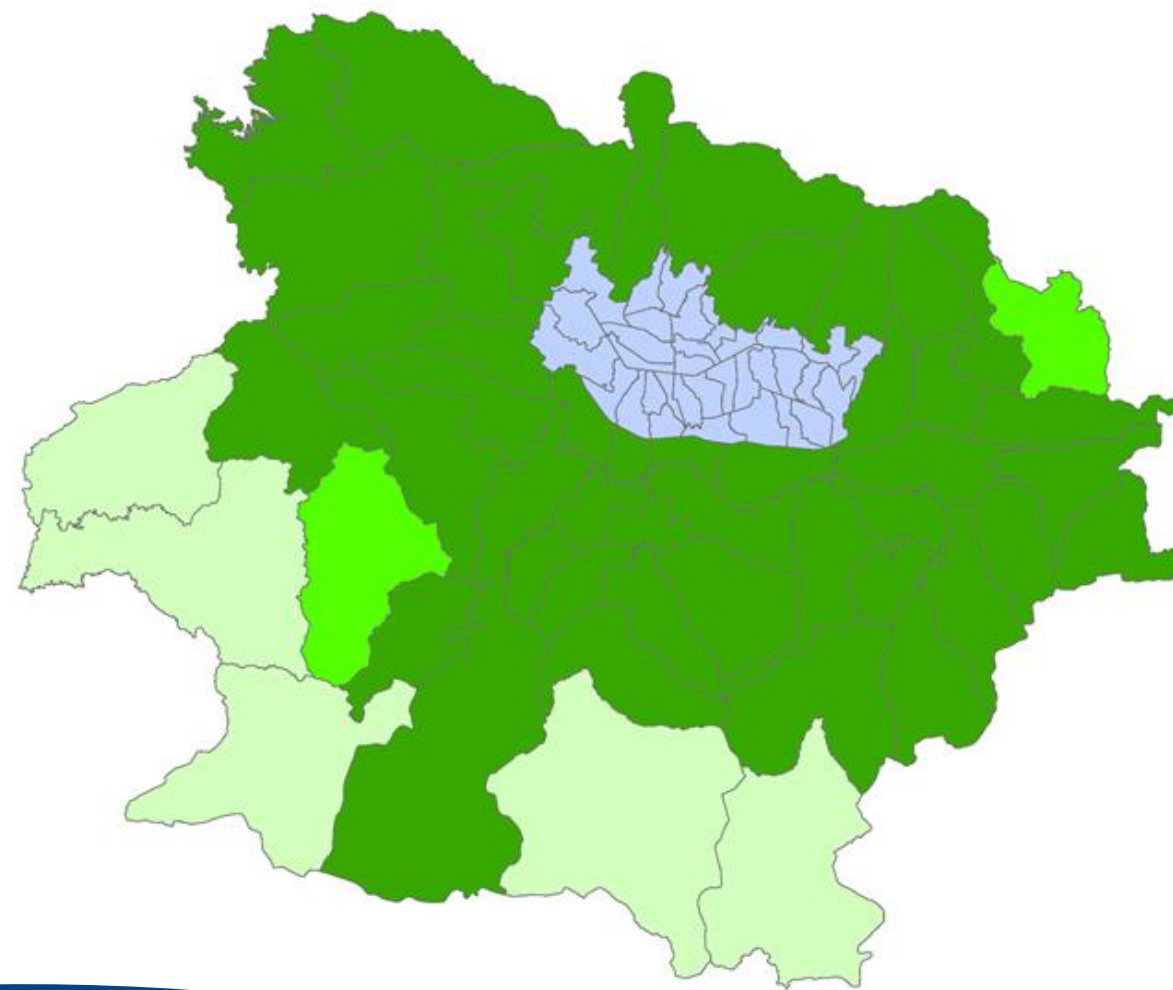
- 5 % of Population
- 1,5 % of Commuters

Using a Travel Diary

(Meili_MSA Application) for validation
(907 volunteers in 1 month)

MA Delineation:

- 45 out of 52 sub-district
- 47 out of 52 sub-district



Central City



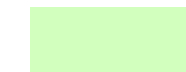
Threshold 5 percent (45 sub district)



+



Threshold 1,5 percent (47 sub district)



Presidential Regulation No. 4 of 2018
(52 sub district)

MPD for MA Delineation on Cekungan Bandung



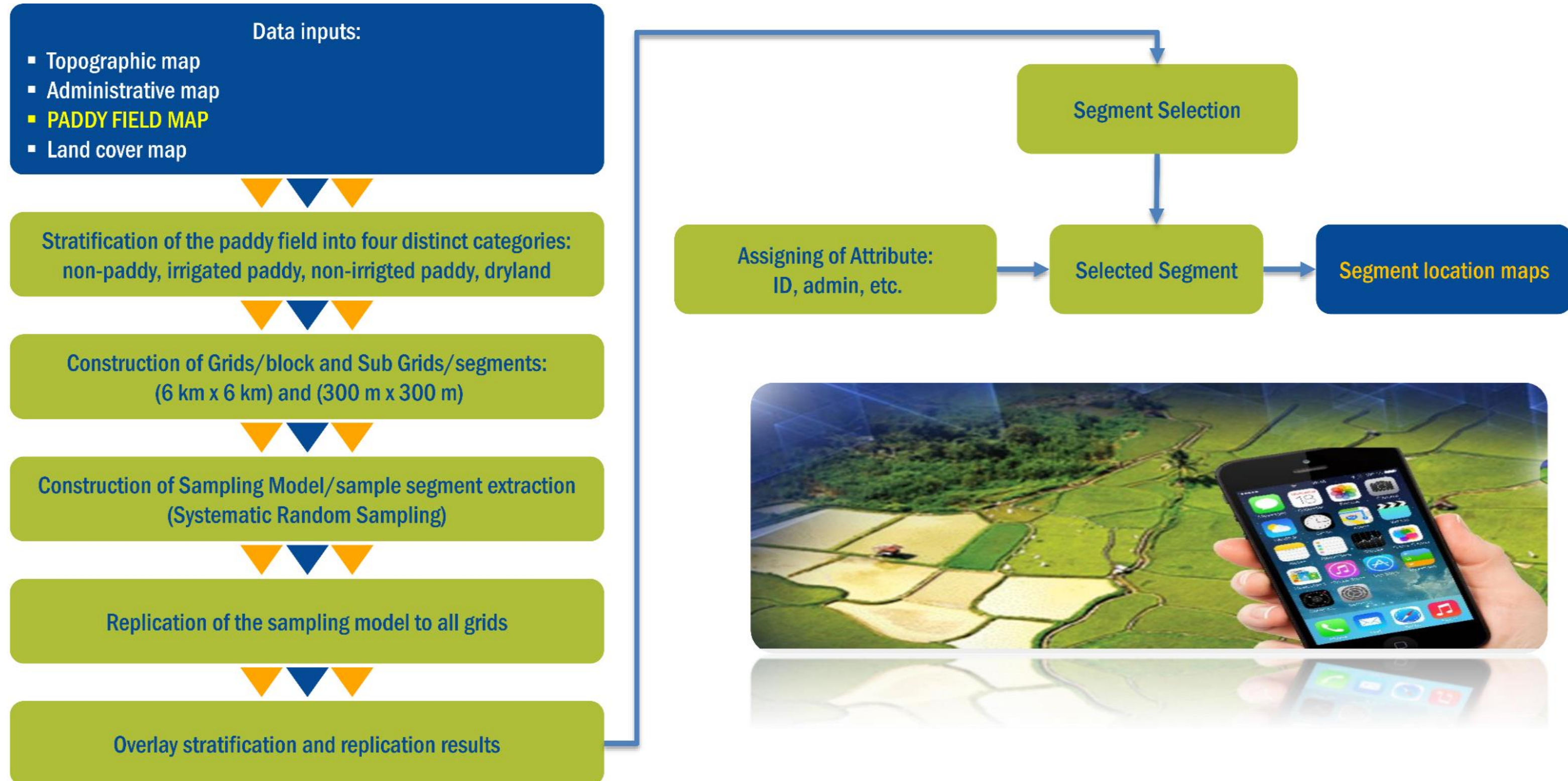
02



The Use of Satellite Images and Android Based Application with Digitalization Data Processing for Area Sampling Frame



Area Sampling Frame

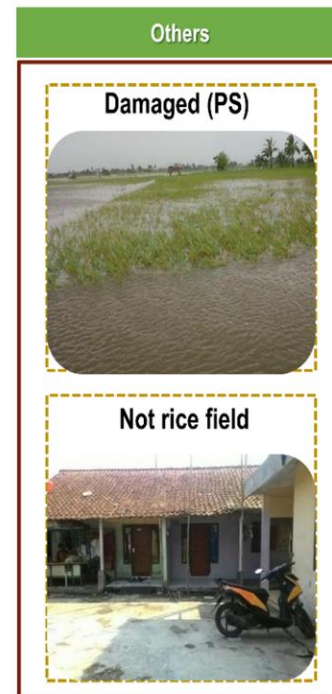
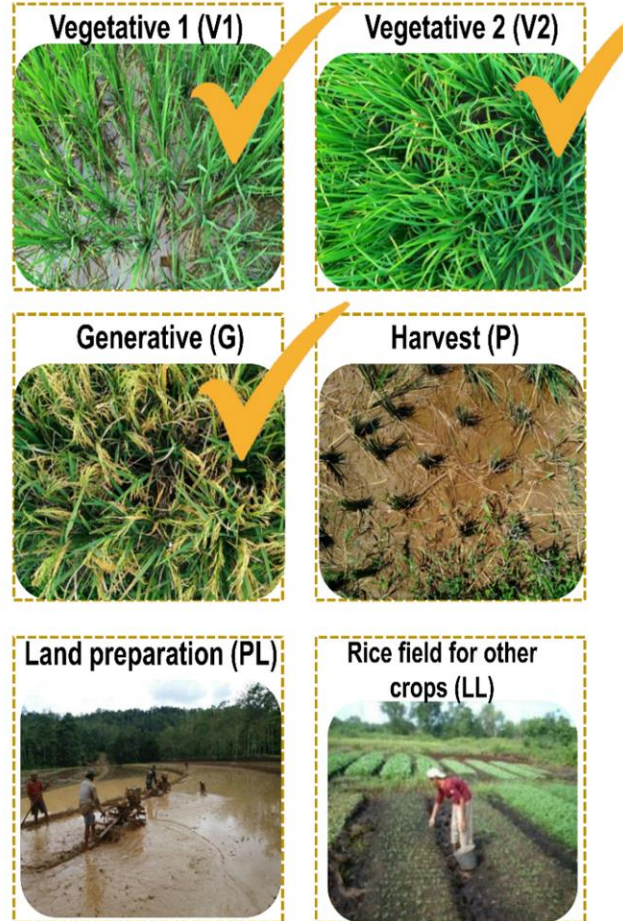


Constructing Sampling Frame And Sample Selection



Capturing with Android Based Application

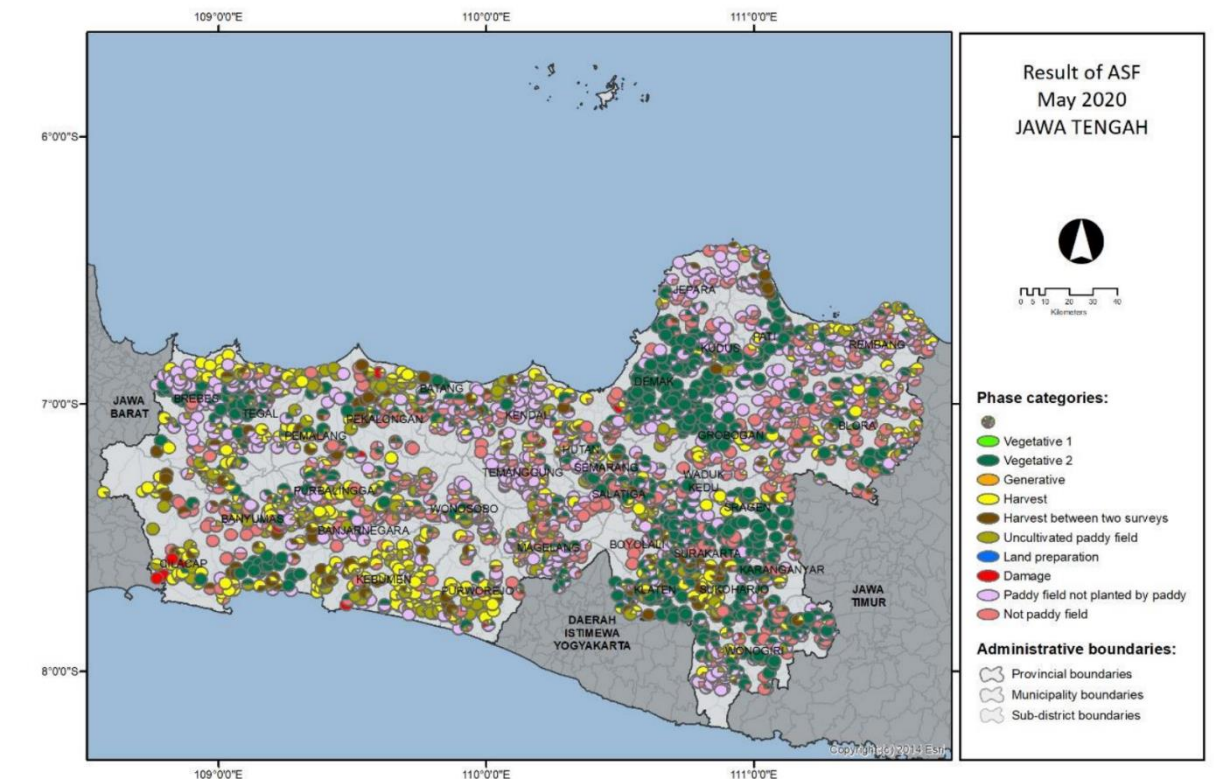
RESULT OF FIELD OBSERVATION FROM ASF SURVEY



Data Results

01

Monthly growth phase monitoring by Province



02

Monthly area estimation for each growth phase by Province

03

Area harvested prediction for the next three months (Potential and Realization of Paddy Harvested Area)

03

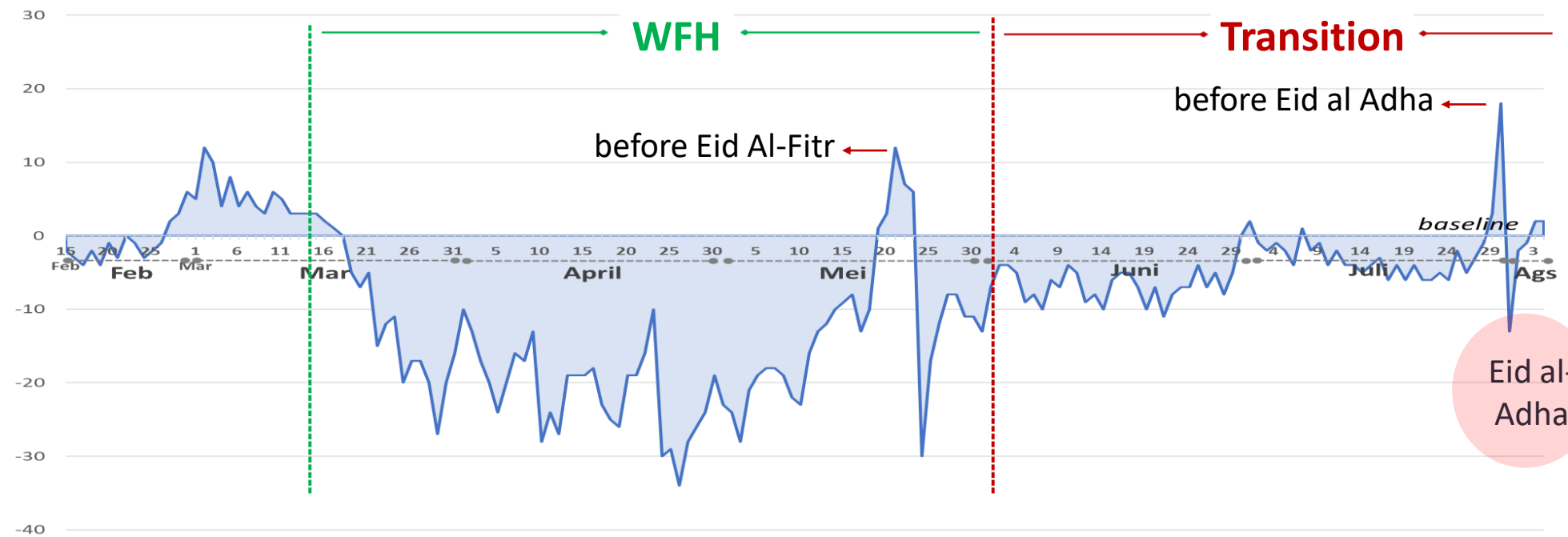


The Use of **Google Mobility** (Impact Covid-19 Pandemic)





Google Mobility Index: Mobility in Grocery Store for Daily Necessities



2 days before Eid al-Adha, activities at the grocery store increased 3% – 8% to the baseline. The similar pattern showed prior to Eid Al-Fitr.

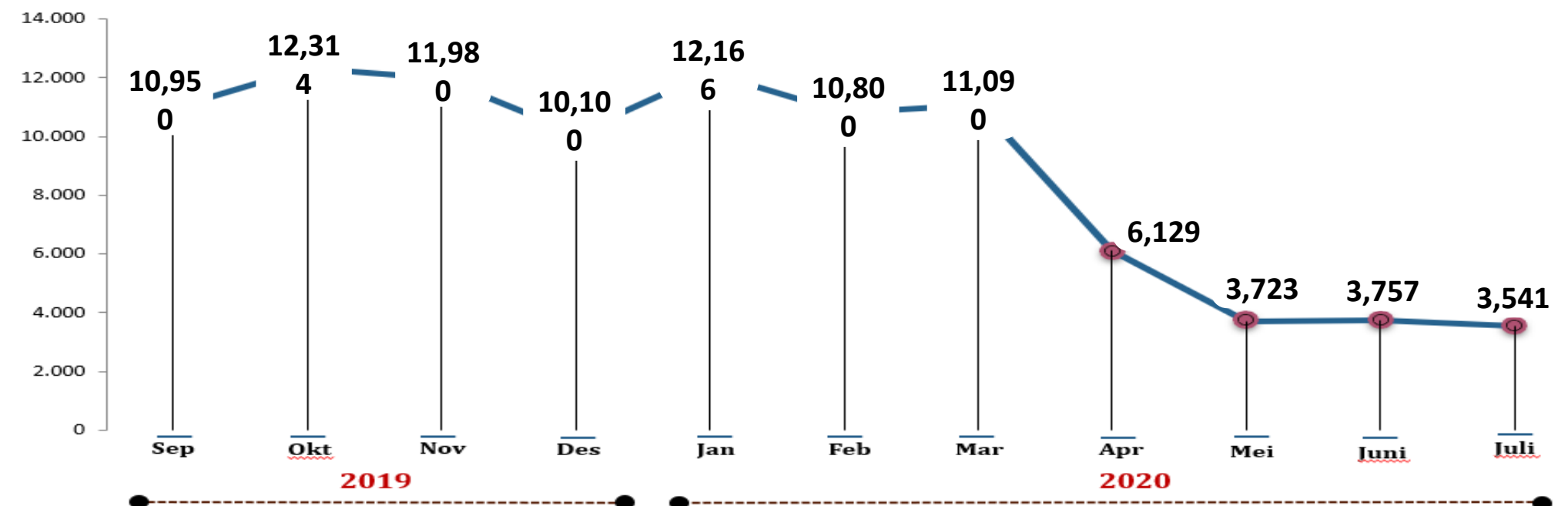
Source: <https://www.google.com/covid19/mobility>



Since the Covid-19 pandemic in March 2020, the number of job vacancies in Indonesia has decreased.

Source: Jobs.id

Job vacancy: Number of Job Vacancies, September 2019 - July 2020





ISSUES AND FEASIBLE SUPPORT

FOR THE DEVELOPMENT OF THE USE OF BIG DATA FOR OFFICIAL STATISTICS



Issue on **Big Data For Official Statistics**

Data Quality

- determine the source and type of the appropriate data to use
- quality assurance
- sustainability of data

Technology

- Integrate multiple data sources
- Handling large dataset
- IT issues

Data Management

- Data storage
- Data processing (machine learning, data mining, data cleaning, etc)

Data Access

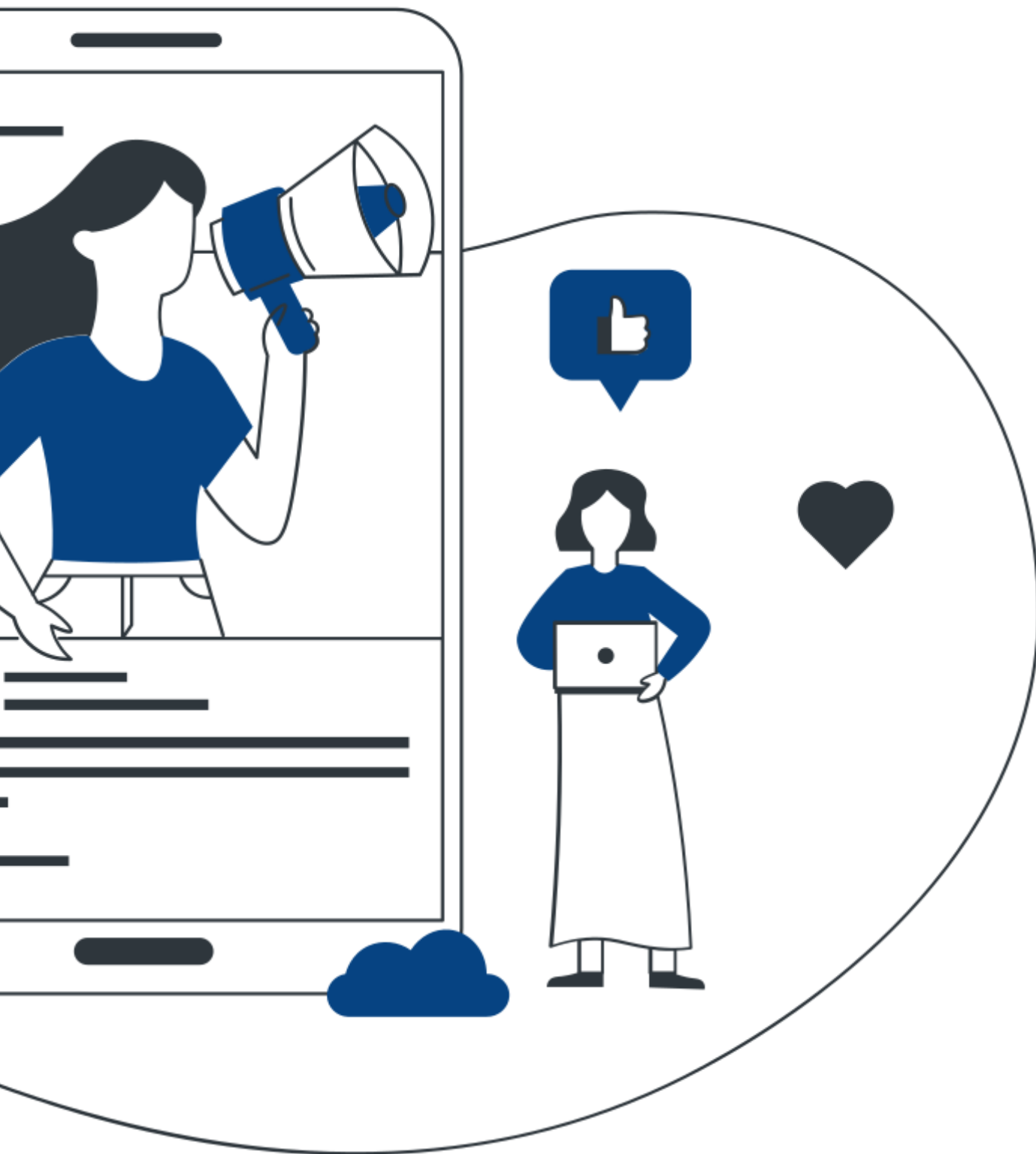
- Legal aspect
- Data security
- Data privacy

Analytical Capability

- Knowledge in statistics (modelling, unstandardized methodology)
- meaningful insight of Big Data

Others

- public trusts



Feasible Support

- Provide strategic vision and direction on the use of Big Data for Official Statistics
- Promote coordination between countries
- Support for skill development and capacity building
- Guidelines for data processing and standardized methodologies in data use
- Providing a platform or environment for the development of big data usage

THANK YOU

“

Big Data offer a number of opportunities for producing statistics, but some big challenges should be overcome

”

Contact us:

Website: bps.go.id

Email: ediawan@bps.go.id