

Experience of Indonesia in Developing Big Data for Official Statistics

Setia Pramana

*SESRIC Webinar Series On
Statistical Experience Sharing*



Thursday, 10 June
2021



BIG DATA

“A paradigm for enabling the collection, storage, management, analysis and visualization, potentially under real-time constraints, of extensive datasets with heterogeneous characteristics.” (International Telecommunication Union, 2015)





TAXONOMY **BIG DATA RESOURCES**



EXHAUST DATA

Passively collected data from people's use of digital services such as mobile phones, financial transactions or web searches.



SENSING DATA

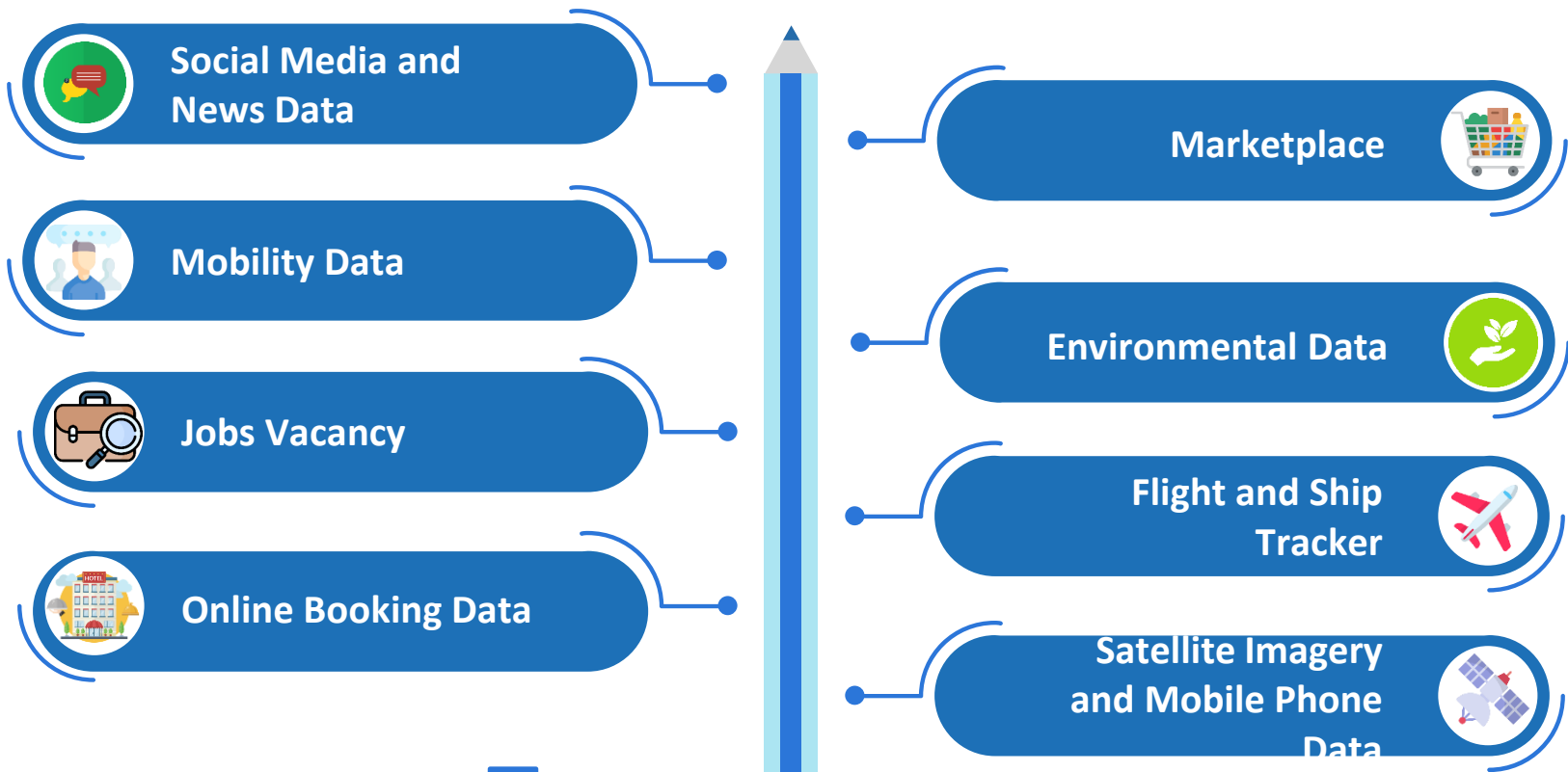
Actively collected data from sensors, e.g. in smart cities or from wearables and also through remote sensing and satellite images.



DIGITAL CONTENT

Open web content actively produced by people such as social media interactions, news articles, blogs or job postings. Unlike exhaust and sensing data is digital content intentionally edited by somebody, i.e. subjective or even deceptive, depending on the intentions of the author.

BIG DATA INITIATIVES IN BPS





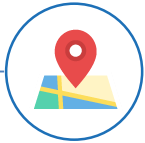
► Mobile Positioning Data (MPD) Use in BPS-Statistics Indonesia



Since 2018 **BPS has had MoU and Contract with MNO (Telkom)** regarding the use of Mobile Positioning Data for official statistics, e.g., Domestic & Outbound Tourism Statistics.



At the beginning, the MPD processing have implemented a hashing procedure to **guarantee the anonymity of the subscribers identification.**

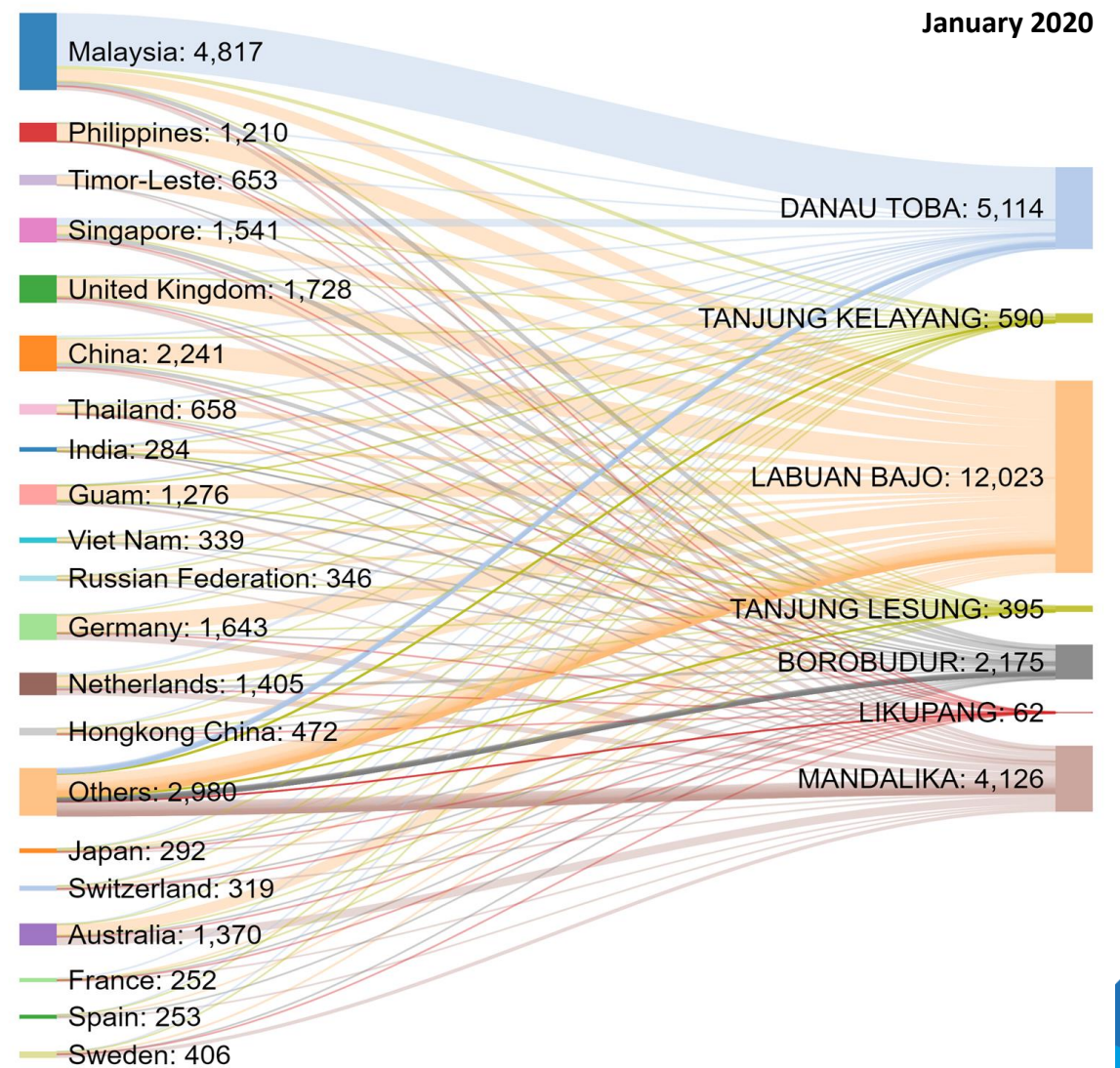
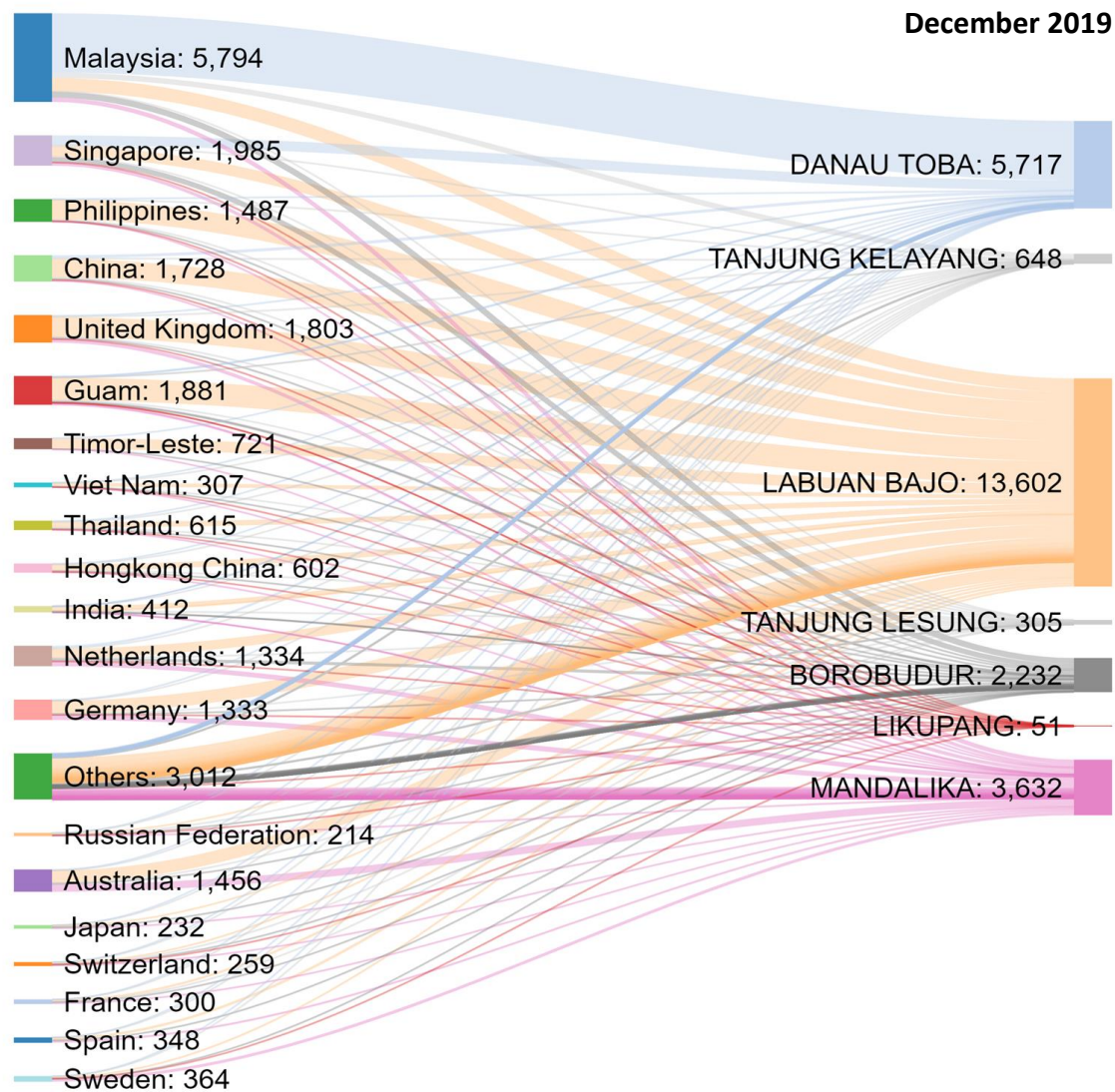


The data of MPD **is only the BTS location** (longitude and latitude), time, and sources (signaling or CDR) which is not the actual location of every subscribers

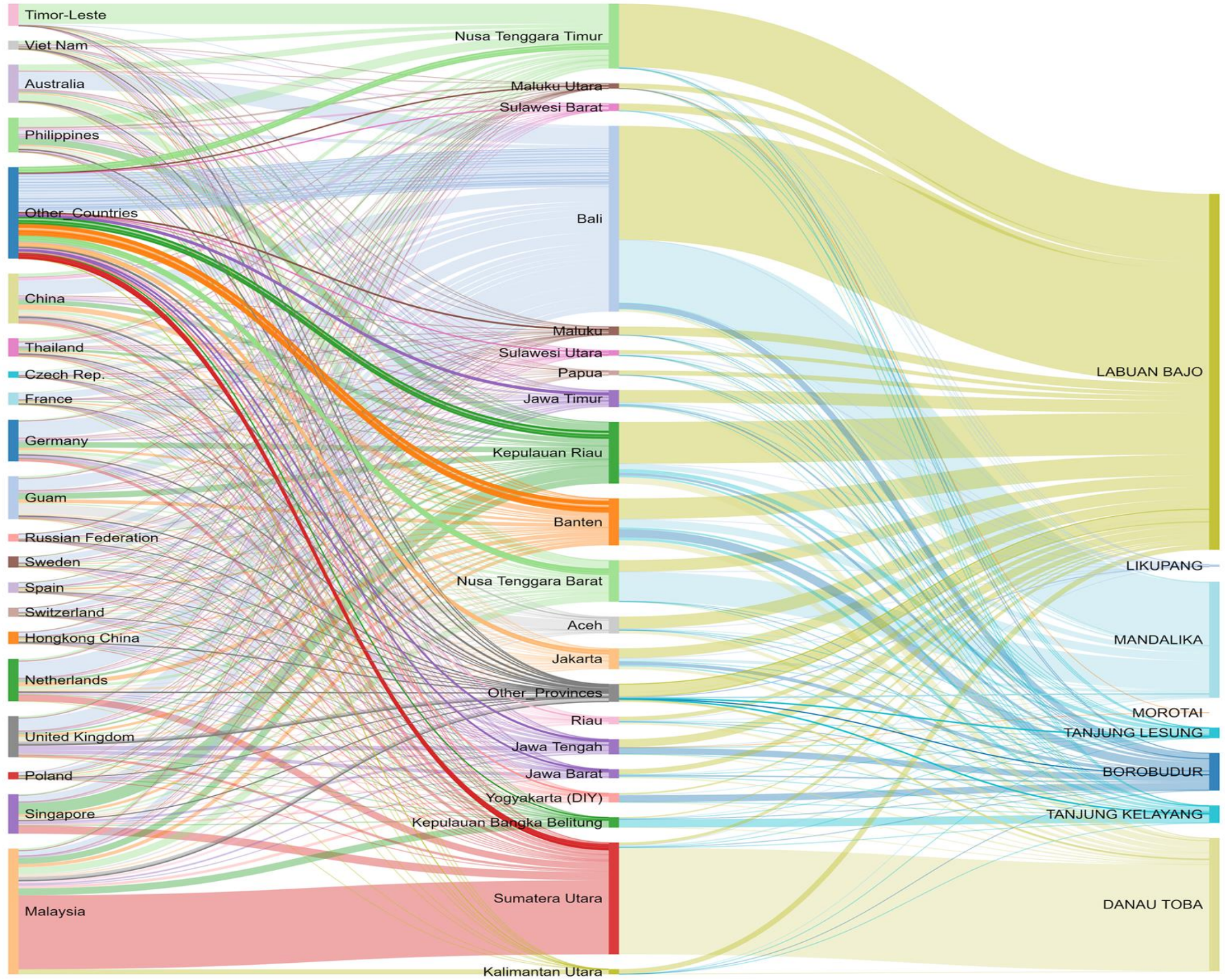


The applications of the MPD data are for the needs of **International Visitors, Domestic Tourists, and Mobility between regions** statistics.

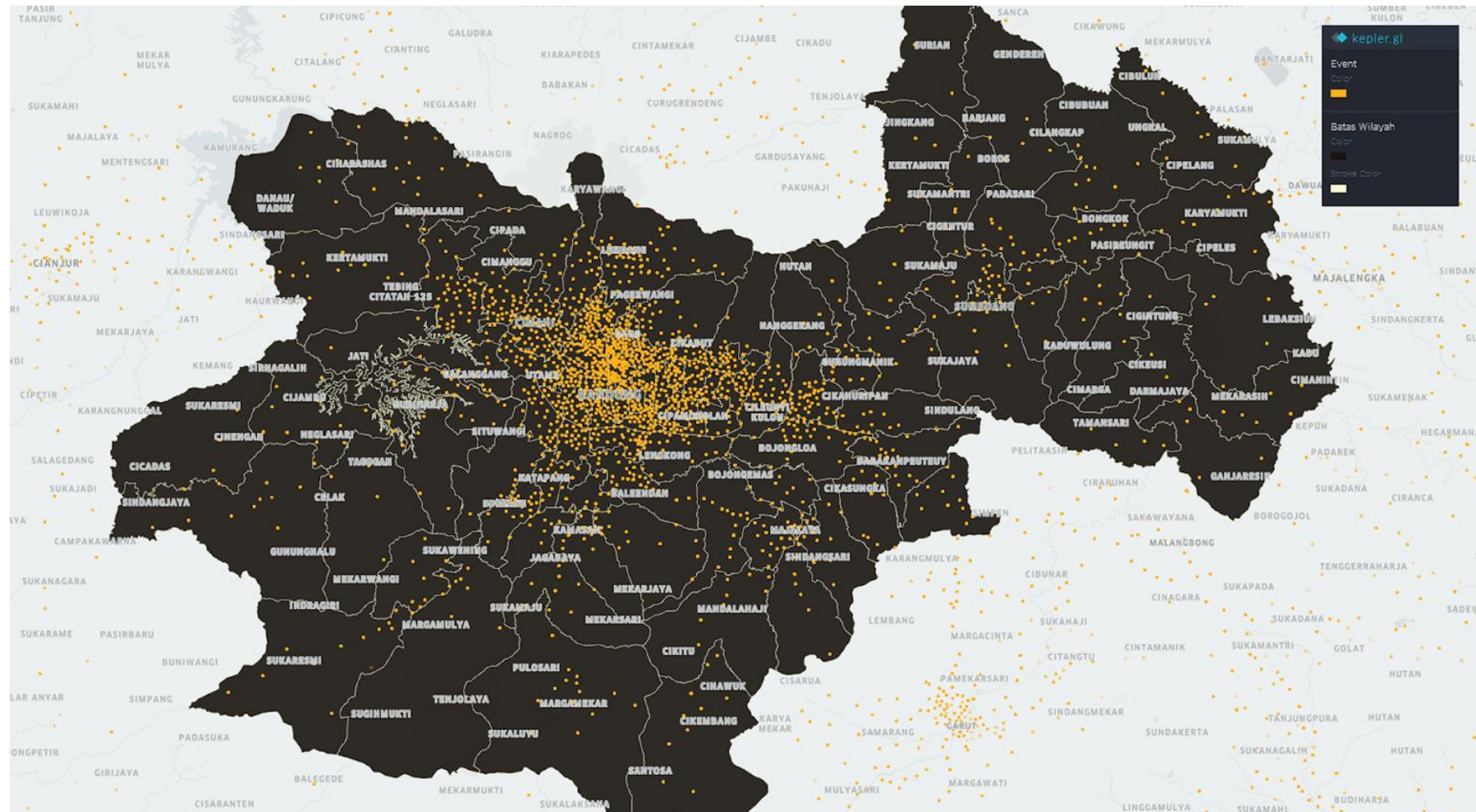
MPD: Foreign Tourists: Monthly Visitors



Countries, Entry points, Entry Destinations

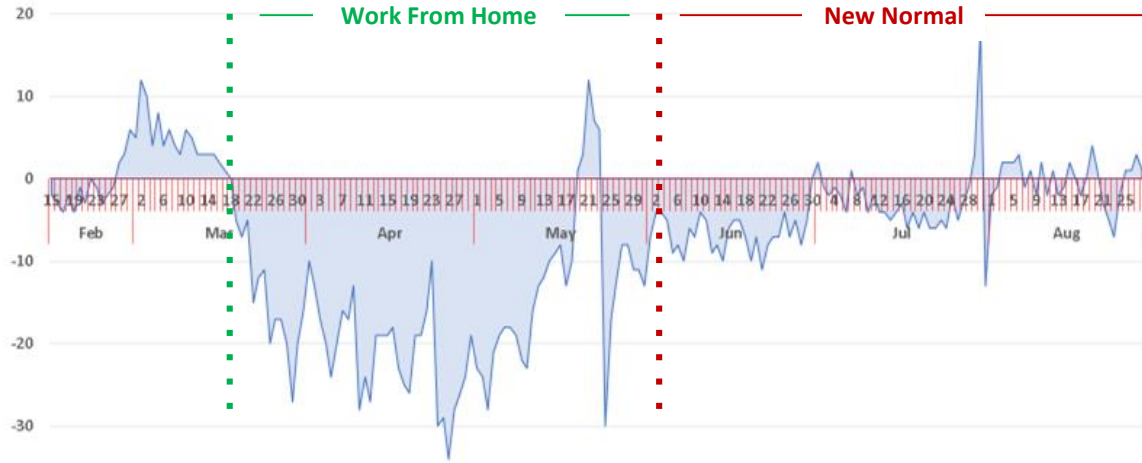


▶ MPD: Metropolitan Statistical Area



BIG DATA IMPLEMENTATION IN PANDEMIC COVID-19 [1]

Google Mobility Index: Mobility in Grocery Store for Daily Necessities



In August 2020, it can be assured that people mobility in grocery store for daily necessities is back to what it was before the pandemic period

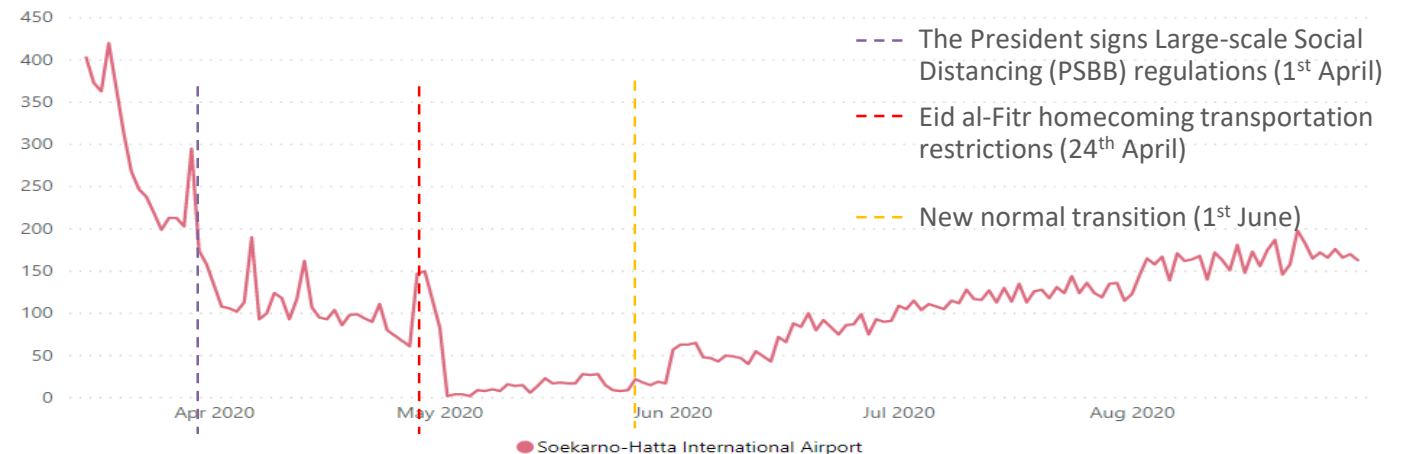
Source: <https://www.google.com/covid19/mobility>

Flight Tracker: Number of Domestic Departures Flight from Jakarta (15th March – 31st August, 2020)



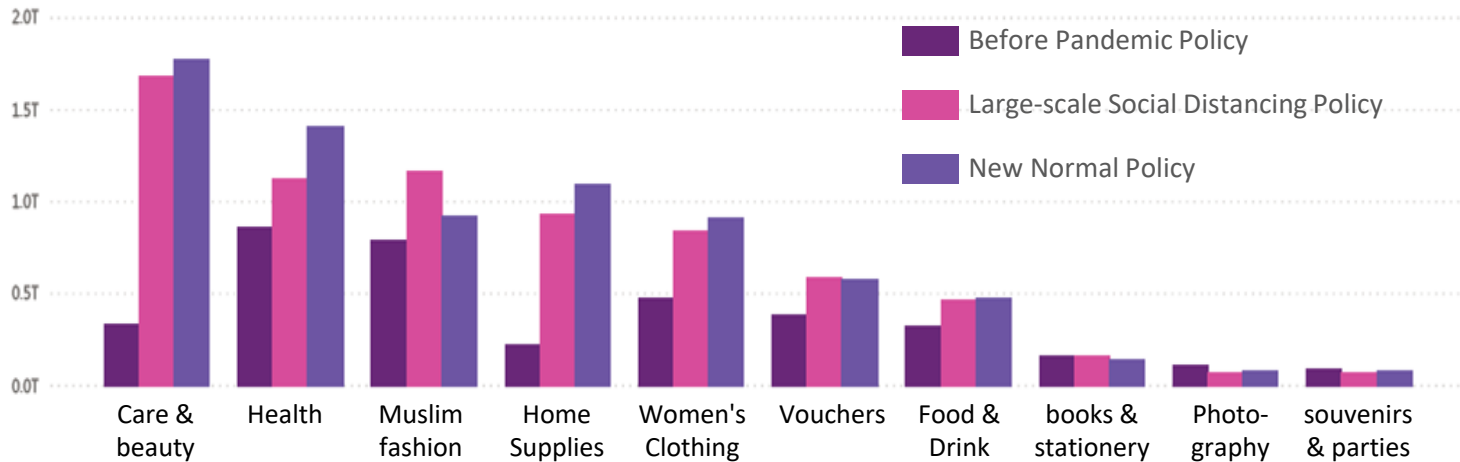
The number of domestic departures flight from Soekarno Hatta Airport increased from 118 flights per day in July 2020 to 158 flights per day in August 2020.

Source: <https://www.flightstats.com/v2>



BIG DATA IMPLEMENTATION IN PANDEMIC COVID-19 [2]

E-commerce Data: Demand Shifting by Product Category in the Online Marketplace



In general, there has been an increase in online shopping activities after large-scale social distancing policy implementation. The care and beauty product experienced the highest increase

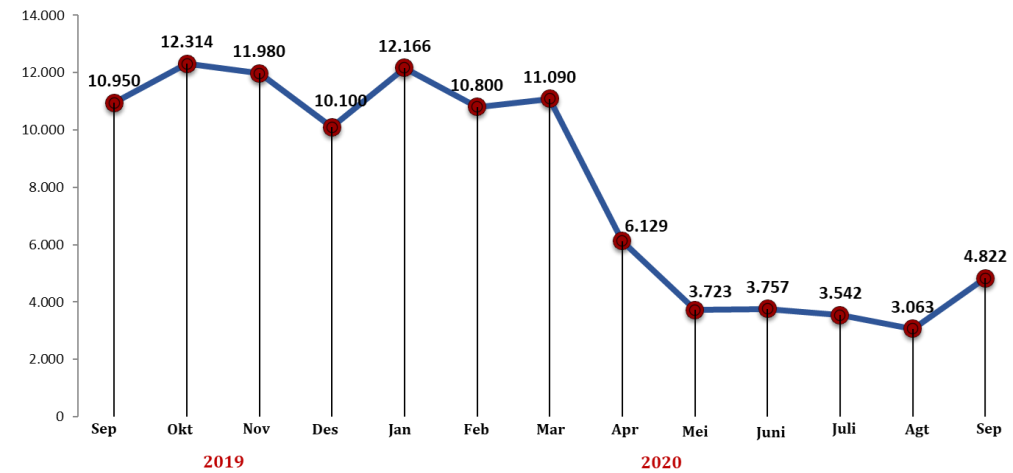
Source: Marketplace Website in Indonesia



The number of job vacancies has been declined since the beginning of COVID-19 pandemic until the new normal period

Source: Jobs Advertisement Website

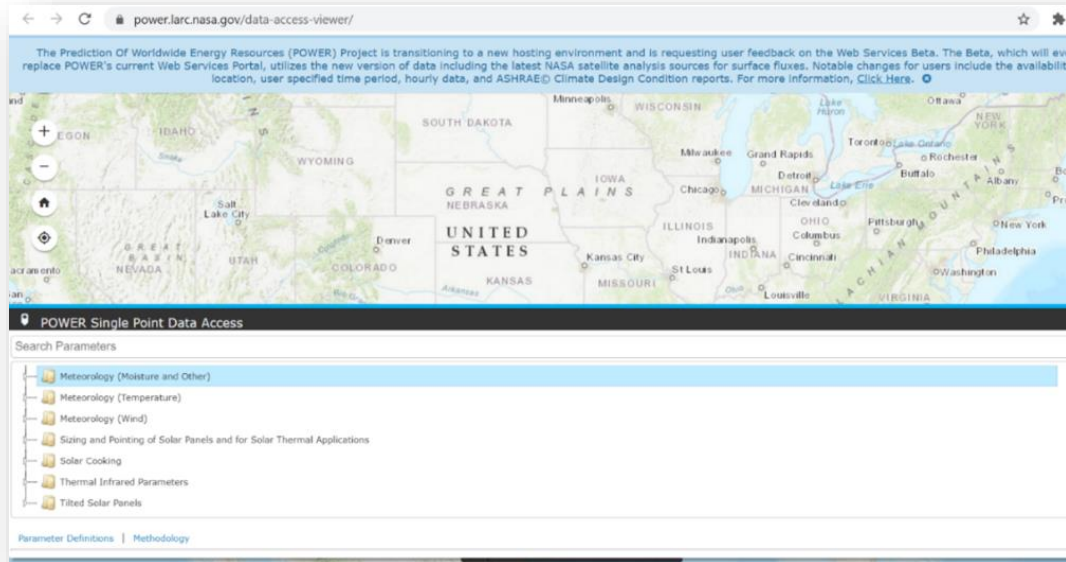
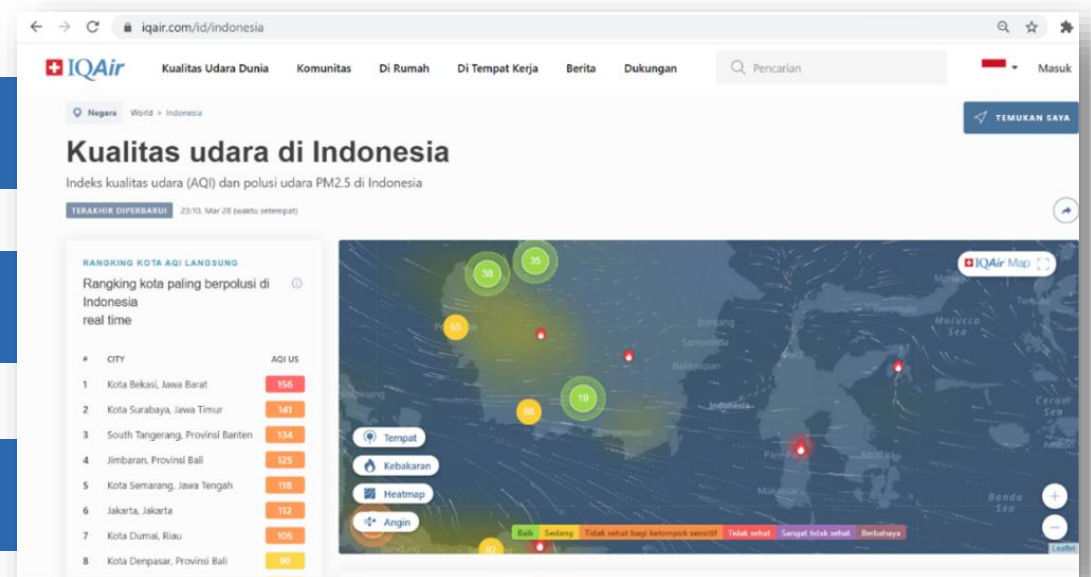
Job vacancy: Number of Job Vacancies, Sept 2019 - Sept 2020



Environmental Related Data

IQair.com

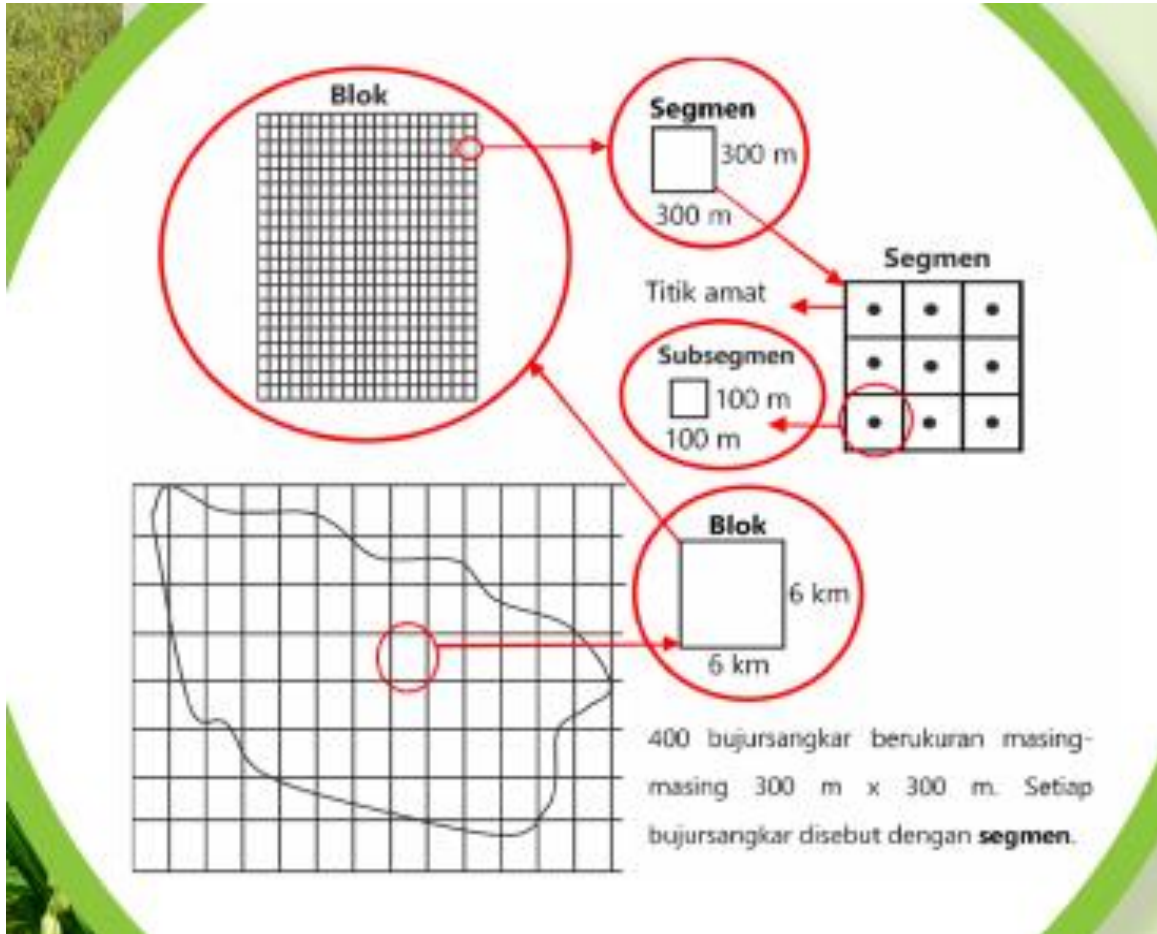
- Variables :
1. Air Quality Index (AQI)
 2. Air temperature
 3. Air pressure
 4. Wind speed
 5. Humidity



power.larc.nasa.gov

- Variabel :
1. Rainfall
 2. Temperature
 3. Humidity
 4. Wind speed
 5. Surface pressure
 6. The temperature of the earth's crust

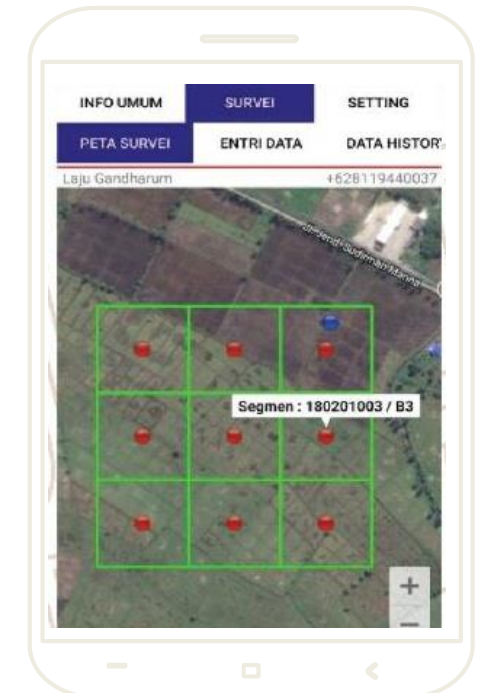
Agriculture Statistics



Selected sample of a segment by SRS to observe its land cover



- Lock the sample coordinate
- Take a picture of the land
- Analyze the sample based on the picture by ML

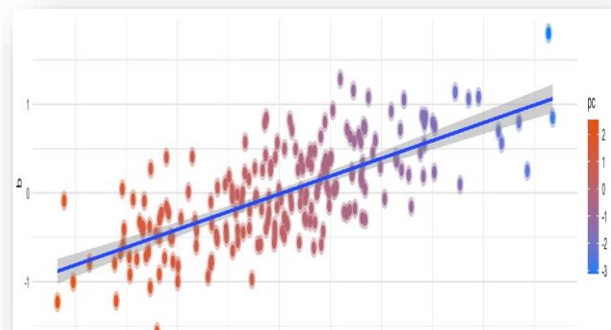
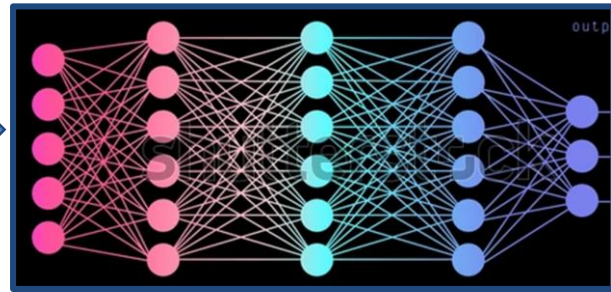


Poverty Statistics – In Progress

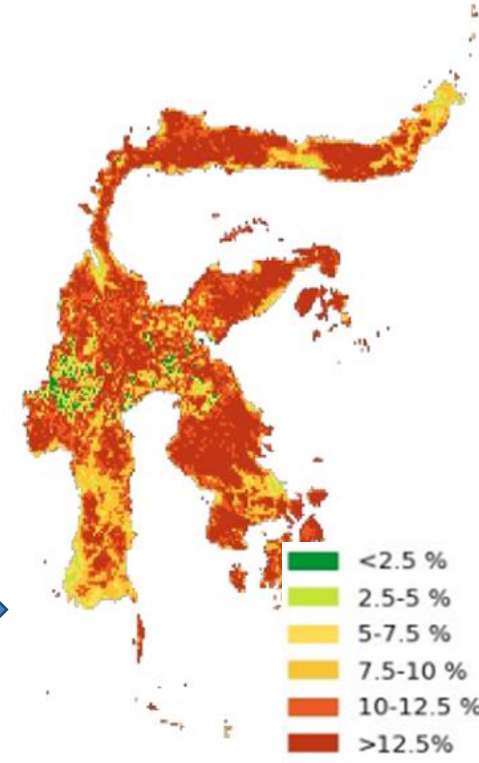
Night Time Light Imagery



Neural Network

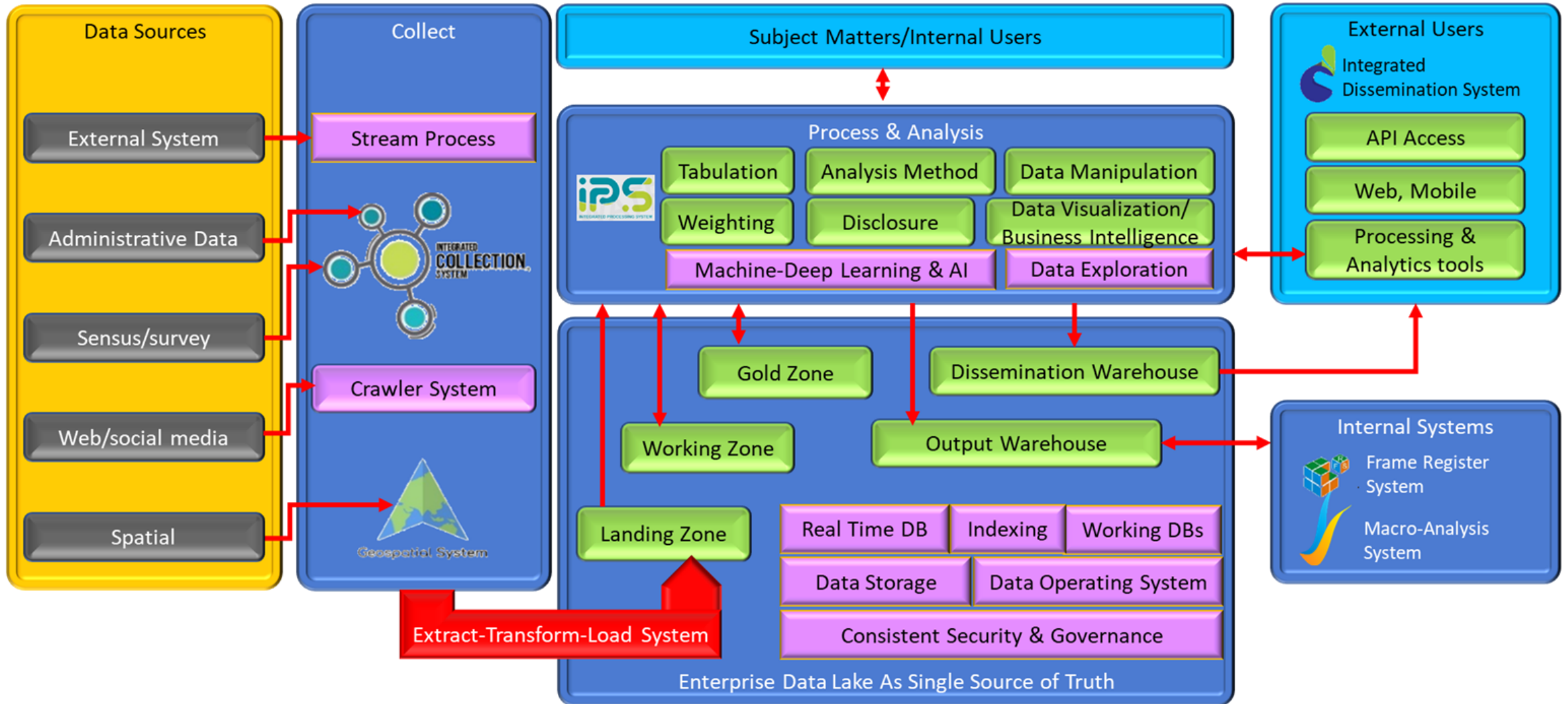


Poverty Map



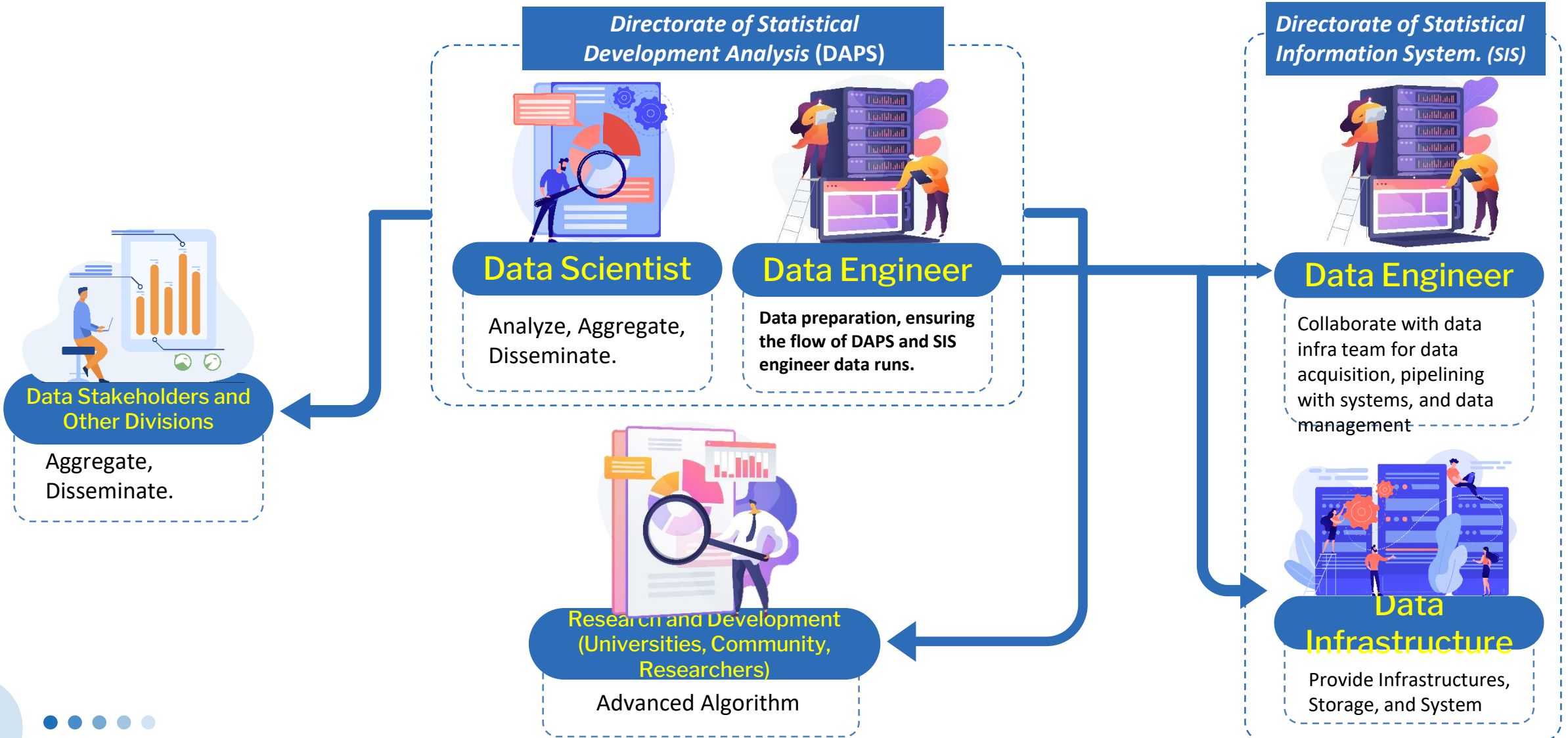
- Use nighttime light imagery
- Measure the intensity of lights to each block (1 km x 1 km) for more precision.
- Apply machine learning to extract features
- Use non-parametric regression to predict poverty levels.

BPS BIG DATA INFRASTRUCTURE





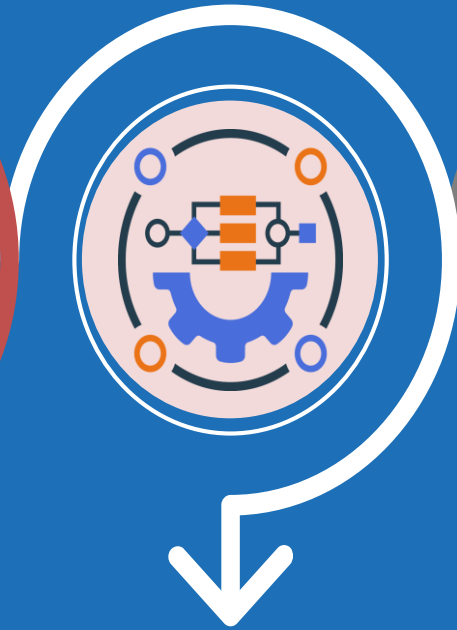
BPS BIG DATA TEAM



CHALLENGES USING BIG DATA



DATA ACQUISITION



STATISTICAL METHODOLOGY



DATA SOURCE QUALITY



**PRIVACY AND
DATA PROTECTION**



**REGULATION ON
NATIONAL STATISTICAL
SYSTEM**



Thank You

www.bps.go.id

