Economic and Social Commission for Western Asia

Big Data: Benefits, Challenges and its Contribution to Statistics on Road Traffic

Mohamad Hossary,
Rim Bahmad

UNITED NATIONS
الاسكوا
ESCWA

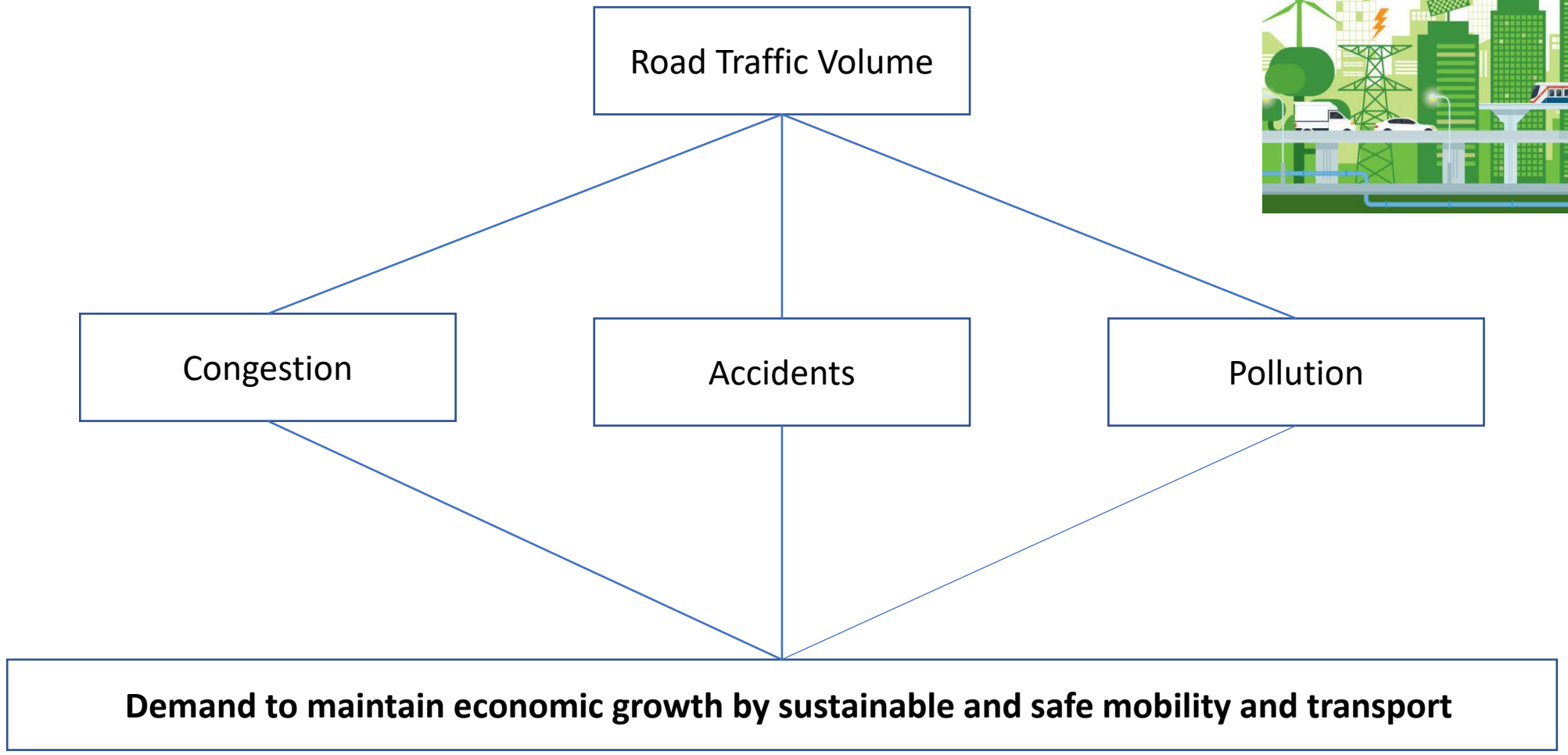
Shared Prosperity **Dignified Life**

SDG targets and indicators related to transport

- (ESCWA) endorsed its interest in collecting transport statistics within the Arab countries since 2019
- 2030 Agenda for ‘Sustainable Development’ adopted by United Nations
- Development Goals (SDGs), which are an urgent call for action by all countries - developed and developing in a global partnership

The aim is to provide by 2030 access to safe, affordable, accessible and sustainable transport systems for all.

How Do Traffic Data Contribute to Achieving SDGs?



Basic Needs for Statistics on Road Traffic

- Analyzing the current transport data and the current transport facilities
- Planning prioritization and project initiation
- Transport policy planning and design
- Establishing appropriate/optimal timings for maintenance interventions and rehabilitation needs of various roads countrywide
- Classifying roads based on their function.
- Assessment of pavement performance
- Road safety evaluation forwarded by adequate measures at specific traffic accident hotspots

Basic Needs for Statistics on Road Traffic (cont'd)

- Analyzing the transport data, evaluating and simulating the traffic data enhance the operational functionality of bus, railway and metro operators.
- Quantifying the amount of fuel used in the upcoming years by providing these data to fuel companies.
- Environmental impact assessment for road traffic in terms of gas emissions, types of vehicles, age of vehicles, fuel type…

All traffic datasets have to be checked based on the quality management rules with a minimum set of meta-information

The data have to be consistent, valid and accurate

Achieving SDGs:

After the implementation of a well-planned, safe and sustainable transport system:

- Accessibility and mobility
- Sustainable cities
- Economy prosperity
- Mitigating pollution and climate change effect

Main Transport Data

- Annual Volume of road traffic (vehicles-kilometer)
  - ➢ Household Survey
  - ➢National Vehicle Register
  - ➢ Odometer readings at roadworthiness tests
- Annual number of accidents

Traditional Data Analysis

1- Data Entry: the data entered in the system should be cleaned of errors before it is summarized and analyzed.

2- Data Analysis:

2.1. Defined vehicle types and kilometers driven,

2.2. Conversion of data in different formats into a common data format,

2.3. Calculation of Traffic Growth Rates of traffic volumes, and production of forecasts based on historical data and growth rates

3- Data Processing and Computer Analysis:

Python, SPSS (IBM), MaxStat, Microsoft Excel, NCSS, STATA, SAS,MATLAB, R (R Foundation for Statistical Computing),etc.

Data Reporting & Presentation

4- Data Reporting: satisfy the specific needs of the users through sending such information in graphic forms to meet the intended reporting format.

5- Data Presentation: Information on traffic data is not always easily accessible. Simple maps or graphs must be produced.

Big Data

- A massive collection of organized and non-organized datasets that are difficult to process using traditional statistical methods due to data complexity.
- Defined as Three V's: high **Volume**, high **Velocity**, and high **Variety** of data that are cost-effective to enhance the decision made by transport planners.
- Provide novel insights through the *relational* qualities of what is recorded (e.g., the ability to link observations of different activities via the common fields of time and space)

Why So Much Data?

In recent years there have been enormous technological advances in the capture and storage of data and providing possibilities to track and triangulate diverse data sets:

- Government transactions (e.g., tax, social security)
- Official registration/licensing
- Commercial transactions by individuals and organizations
- Internet data from search and social networking activities
- Tracking data
- Image data (e.g., aerial/satellite images, land-based video)

All of these are potentially relevant to planning transport systems since they either provide insights into the location, timing and frequency of activities that generate travel (such as employment, shopping or social engagement), or they provide direct evidence of the volume, concentration and direction of person movements or vehicular movements.

Big Data Providers

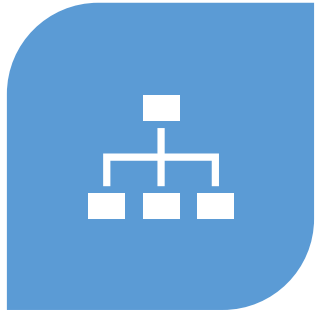
- Social security data to infer trip patterns
- Annual vehicle test data to understand social variations in vehicle use
- Credit card data to reconstruct individual movements
- Public Smart card data with a range of applications in public transport planning.
- Location-based, social media check-in data from services used to estimate travel activity patterns
- Automatic vehicle identification (AVI) technology such as loop detectors, automatic traffic counters to understand complex travel activity patterns.
- Smart phone data and GPS data are most widely exploited for understanding travel/mobility patterns
- CCTV Cameras

What To Do With Big Data?

COLLECT DATASETS

FILTER THE DATA BASED ON RELEVANCY

ESTIMATE RELATIONSHIPS AND CORRELATIONS



ANALYZE, STRUCTURE AND ORGANIZE

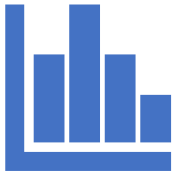
Benefits of Using Big-data

- Better evaluation for the current road traffic condition, and the driving factors affecting it.
- Better decision making for government, planners and business
- Increased planning and operational efficiency
- Aware users of the most appropriate/efficient mode of transport at any given time
- Reducing terrorism and cybercrime through CCTV cameras
- Saving lives by reporting the traffic accidents through all parties using Intelligent Transportation Systems (ITS)

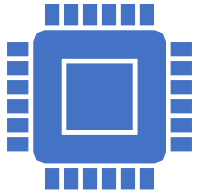
Challenges of Using Big-data

- Data fragmentation among multiple systems
- Breaching privacy and security
- Identification and promoting suitable commercial business models which benefit all parties
- Smaller cities and towns might not be able to see the benefit of Big Data –Capital-centric
- Recruiting skilled workers

Introduction to ESCWA's Initiatives



Working on identifying new technologies and data sources



Actively experimenting with and exploring use of novel technologies and data sources

AIS Data

Google Earth Engine

Web Scraping

Cell Phone

Actively cultivating partnerships for development and implementation

Road Safety, Complementary Data, and Machine Learning Paper Objective

1

Understand severe car crash injuries

2

Demonstrate the use of complementary data sources

3

Demonstrate the use of Machine Learning algorithms

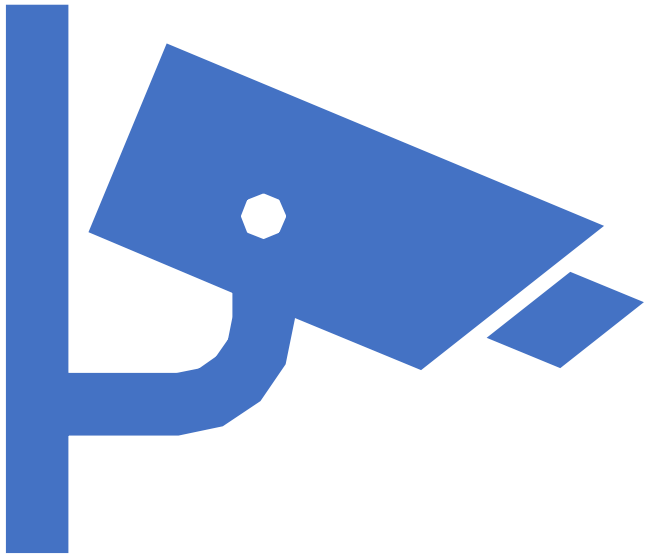
Road Safety, Complementary Data, and Machine Learning Data

- Data from target countries was not acquired due to several challenges
- As a result, we relied on open car crash data from the UK
- Used open street maps data as a complementary source

Traffic Volume

PART OF THE COMPLEMENTARY DATA REVIEWED FOR THE PAPER

KEY FOR DIFFERENT TYPES OF ANALYSIS, INCLUDING ROAD SAFETY



Traffic Volume Alternative Sources

- Use cameras on the roads to estimate traffic volume (allows for vehicle classification as well)
- Purchase GPS data from third parties
- Use average speed as a proxy (can be acquired from private entities or from already available infrastructure) examples:
  - Here Maps
  - Google Maps

Traffic Volume Recommendations

- Road traffic volume is critical for road safety analysis
  - Gives better idea of traffic conditions during crashes
  - Can be linked to crash severity
- Disaggregated data is valuable for analysis
- Complementary data has potential, but needs ground truthing and validation

UNITED NATIONS
الاسكوا
ESCWA

Shared Prosperity Dignified Life

Thank You