Expectations for Data in the Age of Google

THE GOOGLE EFFECT **ON DATA USERS:**

- Expectation for **immediate answers** to detailed questions;
- Looking for a "**fit for use**" level of service;
- Higher willingness to **trade off accuracy for timeliness**;
- Expect **neighborhood level of detail**.

THE GOOGLE EFFECT **ACCELERATED IN ECONOMICS STATISTICS DURING COVID-19**

Demand from users cannot be met using surveys as source data.

NEW CHALLENGE — How to use **BIG DATA** to meet user demand →

An Evolving Statistical Business Process

■ SURVEY — Less timely/Burdensome   ■ BIG DATA — More timely/Less burdensome

Official Statistics

Question → **Design** → **Collect** → Analyze → Answer

FEATURES OF SURVEYS

Recognize a fundamental change in how the way we produce statistics:

- The statistical process through 2000s started with the design;
- Statisticians meticulously designed the process (e.g., sample sizes, test questionnaires, robust collection systems);
- Data was collected (often supported by legal frameworks that compel response) and analyzed, with conclusions drawn;
- Data quality indicators could be readily constructed (co-efficients of variation, response rates, confidence intervals).

VS

Question → **Collect** → **Design** → Process → Answer

FEATURES OF BIG DATA

The world is fundamentally different from what it was just 20 years ago...

- Statistical process often starts with the data;
- The statistician needs to design around the data, the data are no longer a result of the design;
- This introduces a number of new constraints into the work of a statistician;
- Techniques such as record linkage, automatic coding increase importance;
- Quality indicators and quality assessment is more difficult.

Examples of Use of Non-Traditional Data

GOOGLE PLACES API

**Used to measure COVID-19 impact on closure of businesses:**

- **Business operating status:** information from detailed data (name, address, places ID, pricing level, hours of operations.,…).

- **Data collected on 90,000 establishments in 21 cities worldwide, every other week since mid-April 2020** (2,000,000 observations in database).
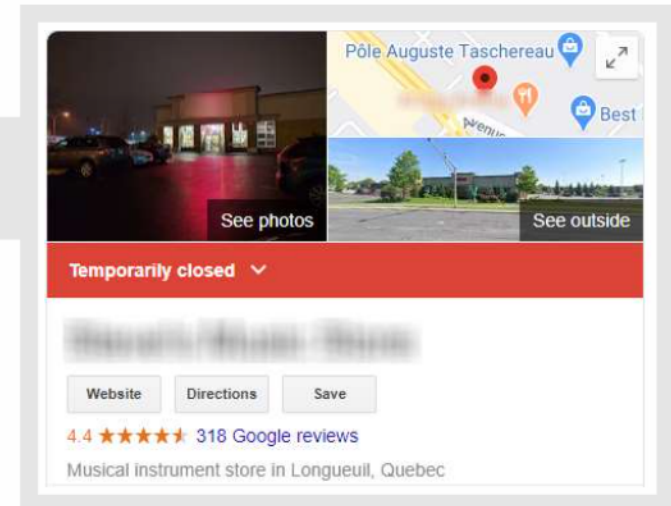
GOOGLE TRENDS

A measure of the interest in a topic relative to all topics over time.

- **Access to anonymized, categorized and aggregated search requests.**

- **Has potential to measure economic activity at high frequency.**
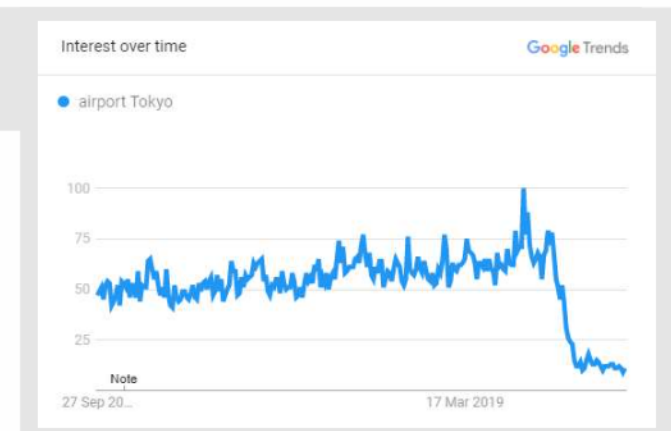
Examples of Use of Non-Traditional Data

WE ARE WORKING WITH THESE DATA SOURCES TO ALIGN THEM WITH STANDARD MACROECONOMIC ACCOUNTING / STATISTICAL CONCEPTS AND CLASSIFICATIONS IN ORDER TO FACILITATE THEIR USE IN NOWCASTING AND THE DEVELOPMENT OF HIGH-FREQUENCY INDICATORS

GOOGLE TRENDS AS AN INDICATOR OF ACTIVITY:

- Google Trends (and Reviews) are used as a **proxy to business activity**, assuming that there is a relationship between changes in interest in a topic(s) and changes in business activity;

- Given the infinite number of possible search terms, Google has developed an algorithm to aggregate searches into **"trend" categories** (e.g., category of "Consumer Electronics");

- The Google Trends categories are then mapped to the relevant International Classification (ISIC), to create **indicators of business activity**.

Google Trends - Selected Retailers

The Home Depot: (Canada) — Canadian Tire: (Canada) — Rona: (Canada) — Walmart: (Canada)
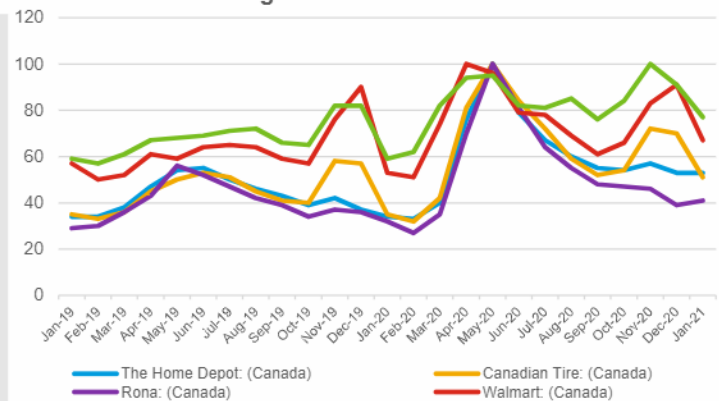
Wholesale and Retail Trade Industry (ISIC Class G)
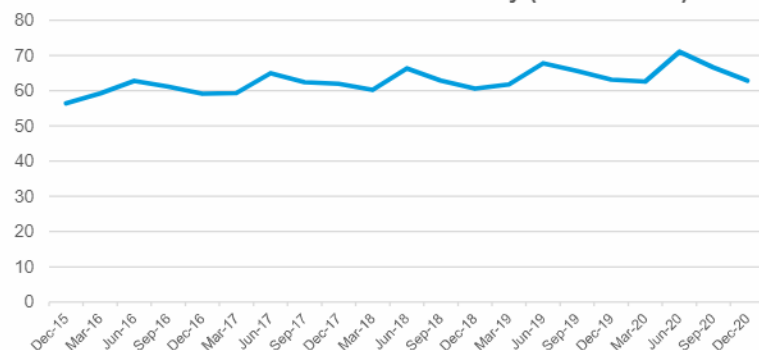
NOVELTY

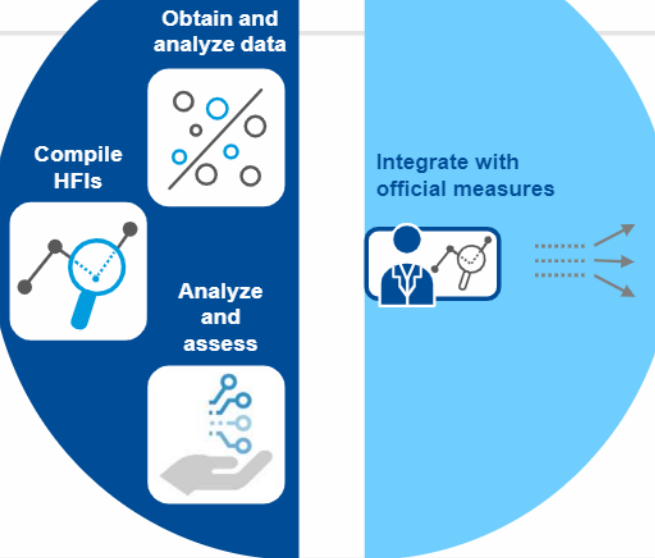# Integrating Big Data with Official Statistics

NATIONAL ACCOUNTS

INTEGRATION
Stage 1   Stage 2

NATIONAL ACCOUNTS NOWCASTS...

QNA Availability

47%   53%

QNA   ANA

Obtain and analyze data

Compile HFIs

Analyze and assess

Integrate with official measures
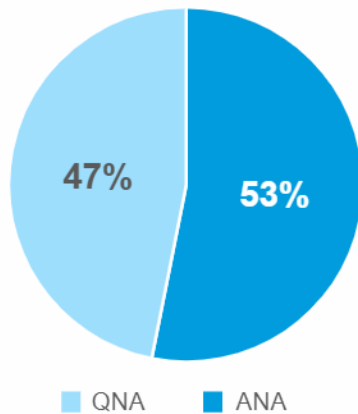
Timelier Data on Economic Activity & Growth

Vietnam
Monthly Exports Using Cargo Data
(Benchmarked to Annual GDP-E)-in national currency

Annualized Monthly Exports (billions)   Annual Exports (NA Basis)
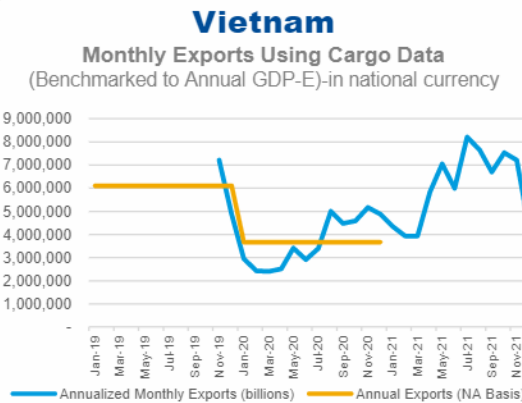
Accurate (and Timely) Nowcasts
of the Second and Third Quarters of 2020

OFFICIAL VS. MODEL ESTIMATES FOR 2020-Q2 AND 2020-Q3

Transportation and Storage (H), Quarterly Rate of Change, Seasonally Adjusted

Legend: ■ 2020-Q2 - Official   ■ 2020-Q2 - Model   ■ 2020-Q3 - Official   ■ 2020-Q3 - Model

Categories: AUSTRALIA, BRAZIL, CANADA, FRANCE, PHILIPPINES, SOUTH AFRICA

Values:
- AUSTRALIA: -21.4, -22.2, 9.0, 10.3
- BRAZIL: -19.0, -18.5, 12.5, 10.5
- CANADA: -26.5, -27.8, 12.1, 26.8
- FRANCE: -23.4, -23.5, 27.5, 28.8
- PHILIPPINES: -59.5, -57.7, 110.0, 80.2
- SOUTH AFRICA: -25.6, -26.8, 15.7, 26.5
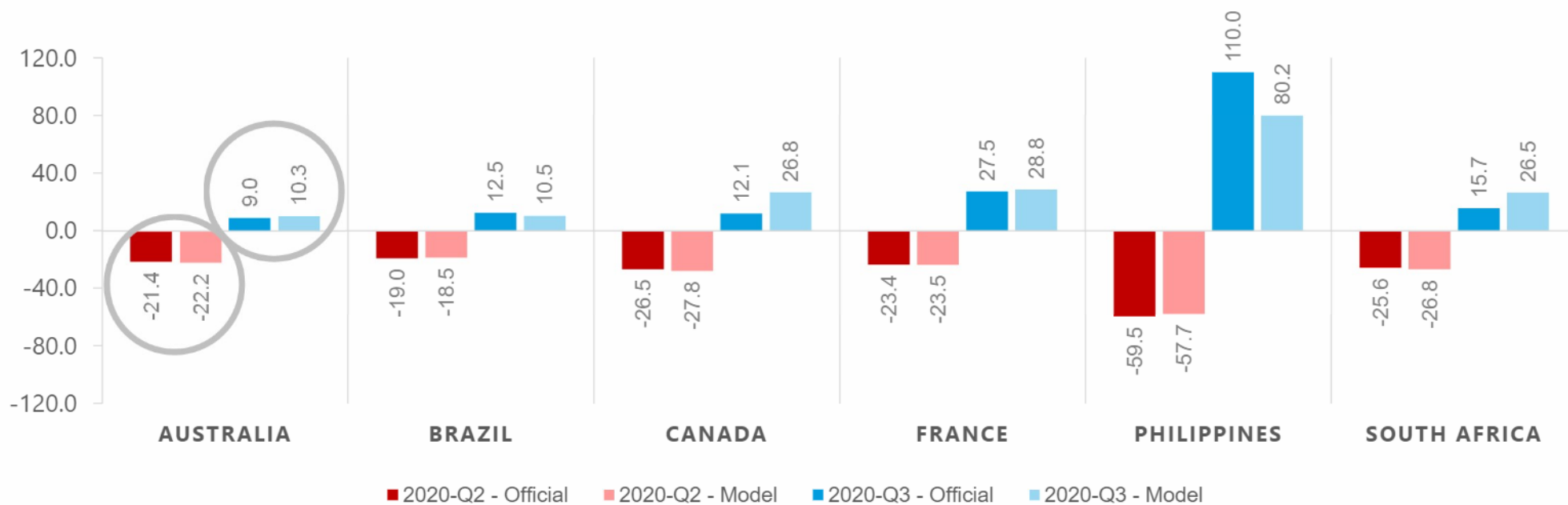
Satellite Data - Nightlights for GDP Estimates

SEVERAL STUDIES USE NIGHTLIGHTS, OBTAINED FROM SATELLITE IMAGERY, TO ESTIMATE GDP OR PROPOSE MEASURES TO IMPROVE OFFICIAL STATISTICS FOR COUNTRIES DISRUPTED BY CONFLICTS AND POLITICAL INSTABILITY

In general, nightlights data are available with a short delay of a few days.

Source: https://www.imf.org/en/Publications/WP/Issues/2019/04/09/Illuminating-Economic-Growth-46670,
https://www.imf.org/en/Publications/fandd/issues/2019/09/satellite-images-at-night-and-economic-growth-yao

Satellite Data - Earth Observation Data

USING SATELLITE IMAGERY DATA TO IMPROVE OFFICIAL STATISTICS ON A WIDE RANGE OF TOPICS INCLUDING AGRICULTURE, CLIMATE, BUSINESS ACTIVITY, AND TRANSPORT

Pilot projects led by UN include crop density, agricultural statistics, land cover and use statistics, urban-rural systems, climate data, and crude oil inventory.

Source: https://cropmonitor.org/
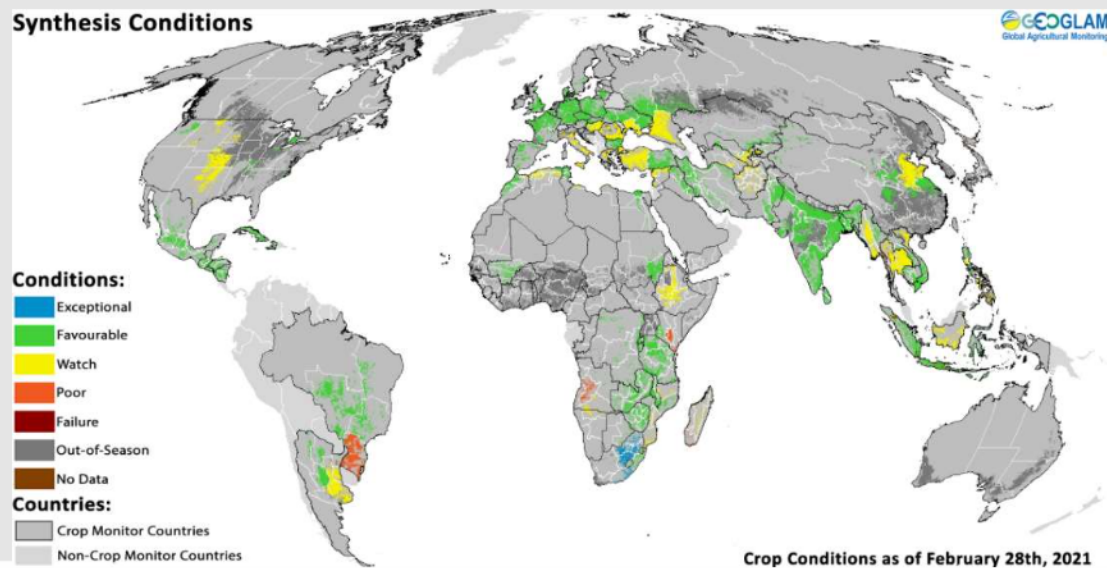
# **Flight Data -** Tourism Activities Estimate

AVAILABLE DATA ON FLIGHTS CAN BE USED TO ESTIMATE TOURISM ACTIVITIES AND AIR TRAVEL IN A TIMELY MANNER

OTHER RELEVANT USE WAS TO TRACK THE TRAVEL DISRUPTIONS DUE TO RESTRICTIONS RELATED TO THE PANDEMIC

Travel disruptions

Travel restrictions due to COVID-19 have caused massive reductions in air travel

(flights volume index, January 1, 2020 = 100)
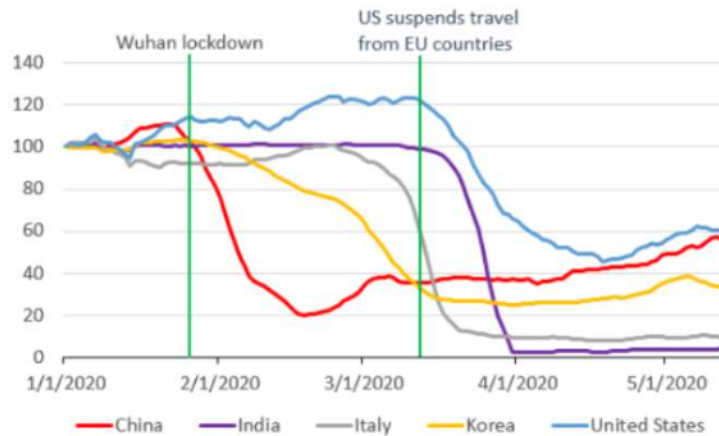
Source: IMF Staff estimates based on Flightradar24 data
Note: Last updated on May 14
https://blogs.imf.org/2020/05/26/keeping-economic-data-flowing-during-covid-19/

Vessel Location Data - Nowcast Trade

AUTOMATIC IDENTIFICATION SYSTEM (AIS) DATA INCLUDE INFORMATION ON LOCATION, SPEED, AND STATUS OF VESSELS (E.G., WEEKLY PORT CALLS AND TRADE VOLUME)

One possible use of data is to nowcast international trade of goods.

A Snapshot of Global Vessel Traffic Based AIS Data

Source: MarineTraffic.
Note: Different types of vessels are shown in different colors.

Scanner Data for Price Statistics Compilation

SCANNER DATA FROM SUPERMARKET CHAINS AND OTHER RETAILERS, AS WELL AS ONLINE PRICES OBTAINED FROM WEB SCRAPING, CAN BE USED TO COMPILE PRICE INDICES

One example is the recent technical assistance mission to the Republic of Kazakhstan, with the objective of assisting the authorities in continuing the CPI modernization, aiming at introducing scanner data in the CPI as a new data source by January 2024.

Combining Different Data Sources
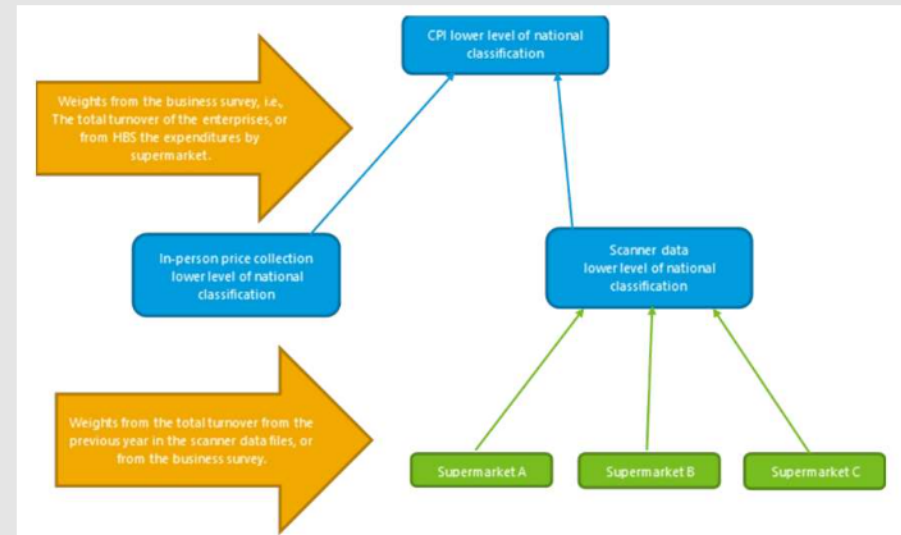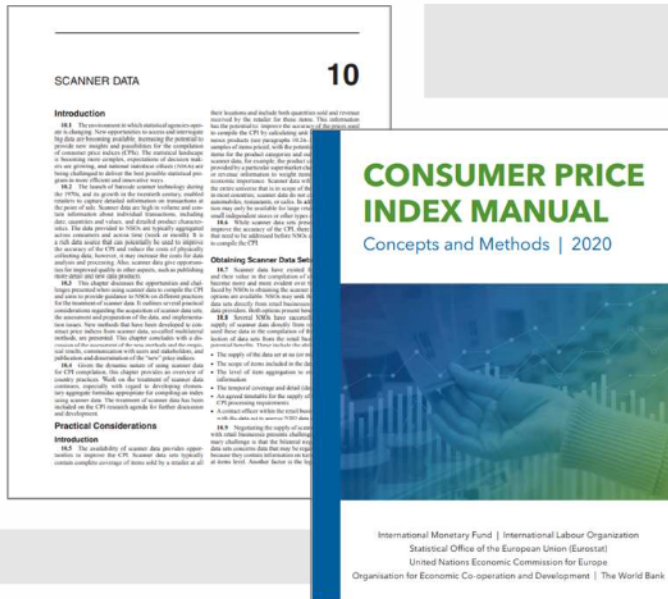
Global Data Sets to Develop HFIs and MIEG

GROWING SET OF OPEN DATA WITH NEAR GLOBAL COVERAGE IS AVAILABLE TO NATIONAL STATISTICAL INSTITUTES

These data can be used to temporally disaggregate annual benchmarks and extrapolate estimates to current period in cases where annual benchmarks are not available.

2019	2020	2021	2022	Global/National Data Set
797,620,028	844,453,296			Satellite Data
2,744,204,956	2,337,928,312			Night lights, Google Trends, Mobility
1,454,225,150	1,169,593,077			Satellite Data
239,337,859	211,197,190			Home sales
1,276,588,357	1,076,901,614			Google Trends, Mobility Data
554,850,839	510,457,680			Flight / Vessel Traffic Data
4,840,799,597	4,712,247,063			Flight / Vessel Traffic, Mobility Data
10,453,401,639	9,693,185,161			

Challenge for Statisticians

How to transform **BIG DATA** into meaningful, quality statistics that can be replicated in order to be produced on an ongoing basis?

BIG DATA

INNOVATION IN

WORKING WITH DATA SUPPLIERS

Cooperation and negotiation replacing laws and penalties:

- Work with private providers to have them sort data according to statistical classifications (for example, price data by sectors)
- Development Partner initiative: IMF and other international organizations working jointly with private sector firms to gain access to data – gain access to data for research.

HR SKILLS

Re-tool knowledge

- Increased need for data analytics skills to enhance interaction among IT, statistics, and businesses, improving efficiency and performance by discovering patterns in data.
- Although still relevant, less need for traditional statistical, sampling and data collection skills and techniques.
- International Organizations to develop methodologies to facilitate the use of Big Data by member economies.

IT

Software, IT

- Increased use of data analysis and modelling techniques, that can be initially based on readily available software and scripts, to be adapted for specific needs
- Although still relevant, less need for in-house development of specific software for data collection and processing.

GOVERNANCE WITH COUNTRIES

Establish statistical coordination groups (ministry of agriculture, business registry, tax authority, …) so that public data is accessible for statistics – often gains momentum as part of nowcasting initiatives.

Benefits / Expected Outcomes

- **Nowcast** of traditional economic indicators

- **Increase frequency**
  (e.g., from quarterly to monthly GDP)

- **Examine evolving structural changes in real time**
  (e.g., business opening, closing)

- **Assess pandemic impact by business types/geographic locations**

Thank you