STATISTICS

New technologies in real sector statistics

SESSION 2, NOVEMBER 9, 2022

Ms. Margarida Martins and Mr. Barra Casey
Real Sector Division, IMF Statistics Department

Key message

- New technologies have changed user expectations for official statistics
- Countries need to move with technology and respond to the challenges presented

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 meet user demand**

New technology leads to new data sources

■ **SURVEY** – Less timely/Burdensome ■ **BIG DATA** – More timely/Less burdensome

Official Statistics

SURVEYS

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

- The statistical process start 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 constructured (co-efficients of variation, response rates, confidence intervals).

VS

BIG DATA

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

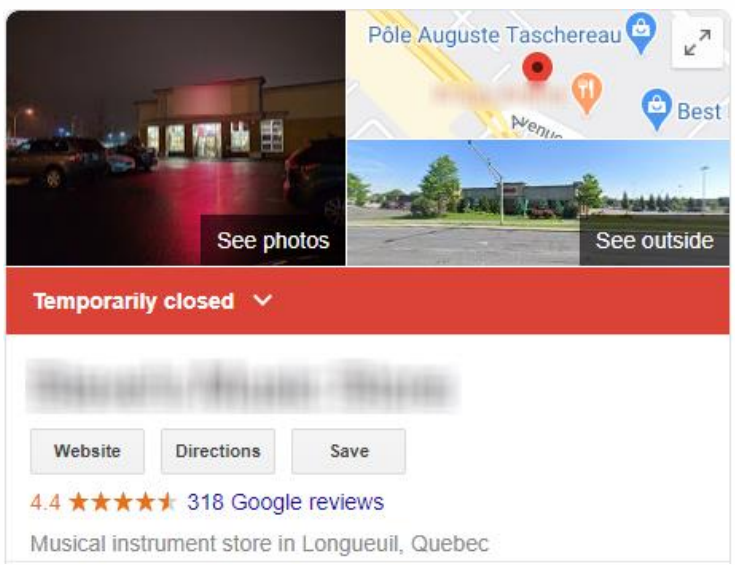
- Statistical process often starts with the data;
- The statistician needs to design around the data, the data are no longer a result of the design;
- This introduces a number of new constraints into the work of a statistician;
- Techniques such as record linkage, automatic coding increase importance;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

Examples of new data sources

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.

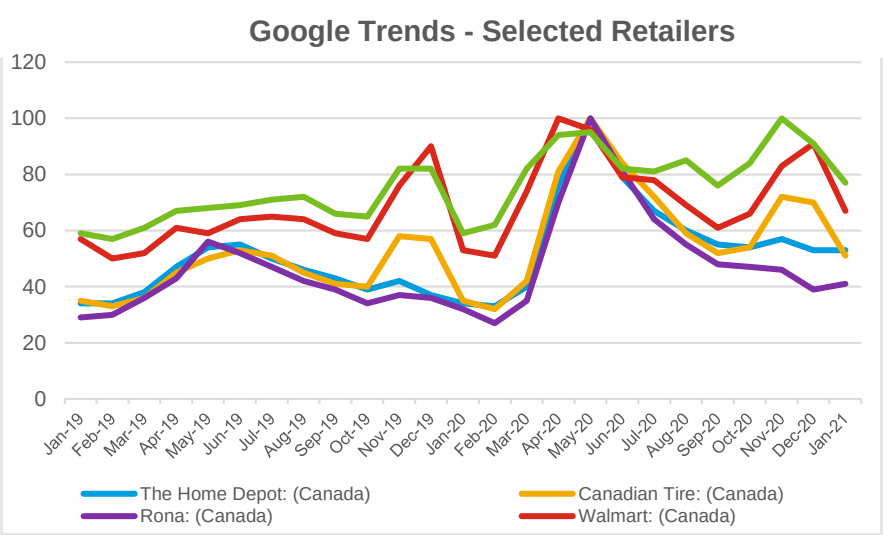
- **Access to anonymized, categorized and aggregated search requests.**
- **Has potential to measure economic activity at high frequency.**

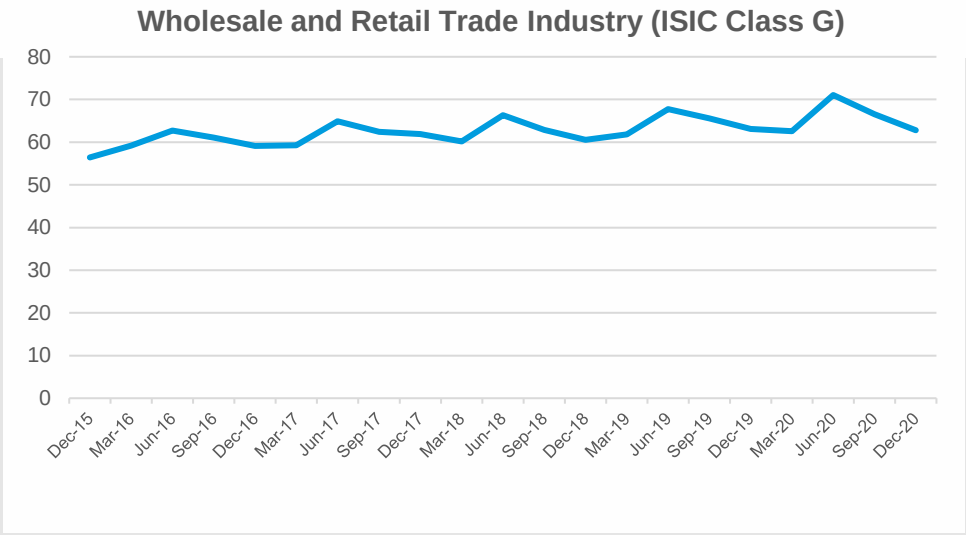
Examples of new data sources

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**.



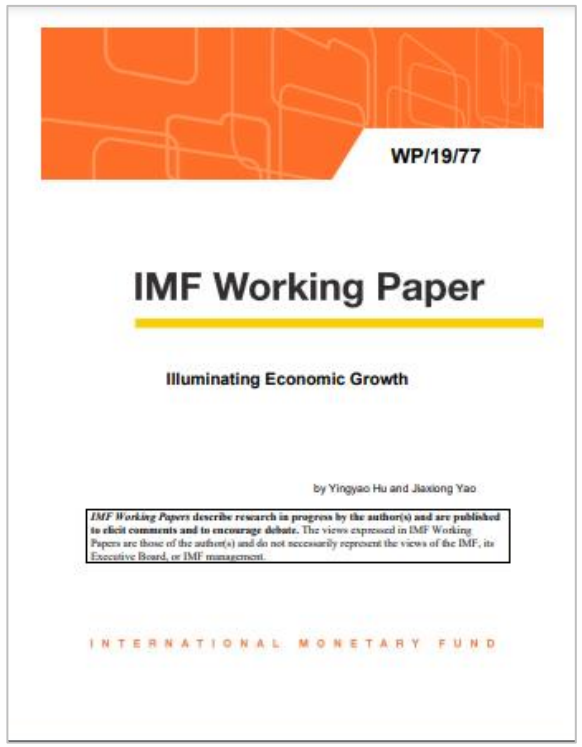


NOVELTY

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.


WP/19/77

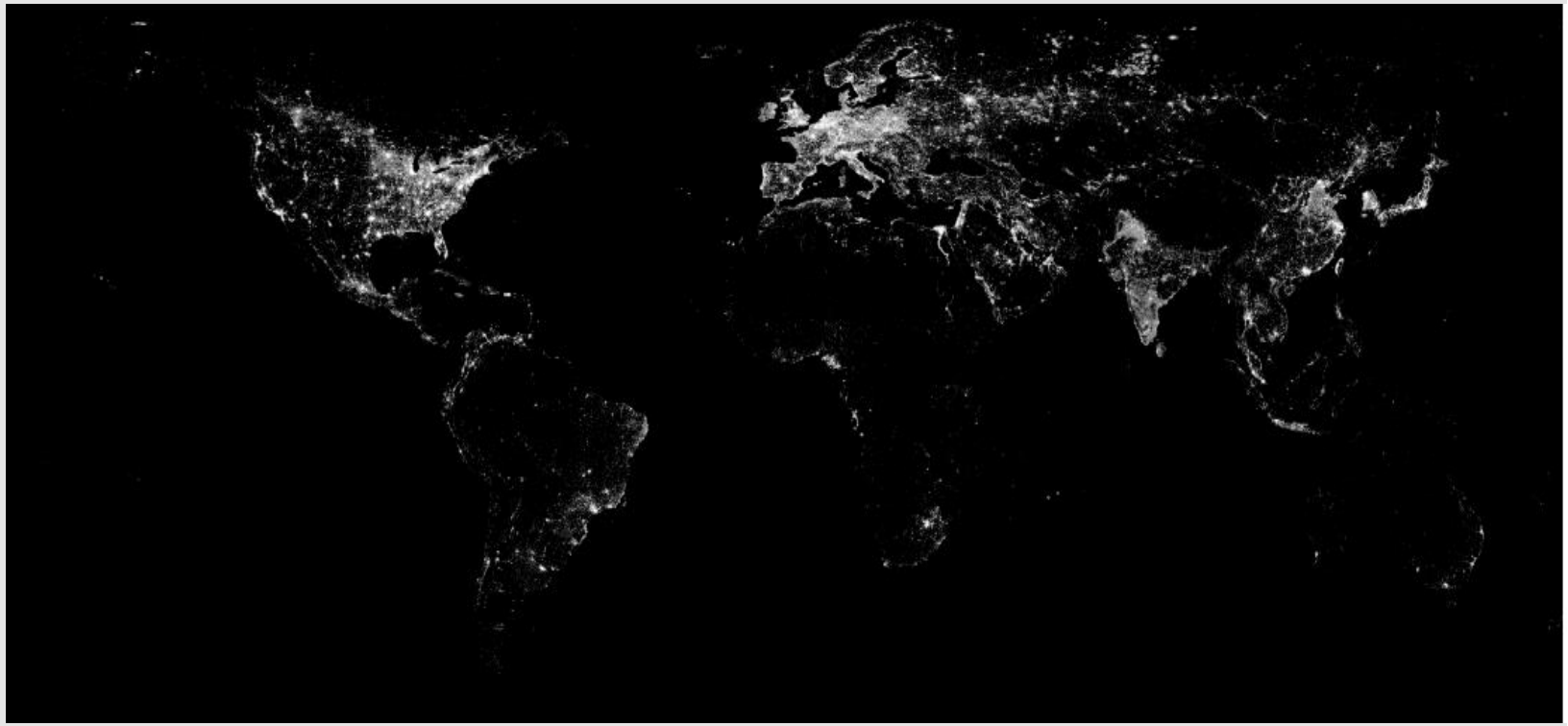
IMF Working Paper

Illuminating Economic Growth

by Yingyao Hu and Jiaxiong Yao

IMF Working Papers describe research in progress by the author(s) and are published to elicit comments and to encourage debate. The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

INTERNATIONAL MONETARY FUND

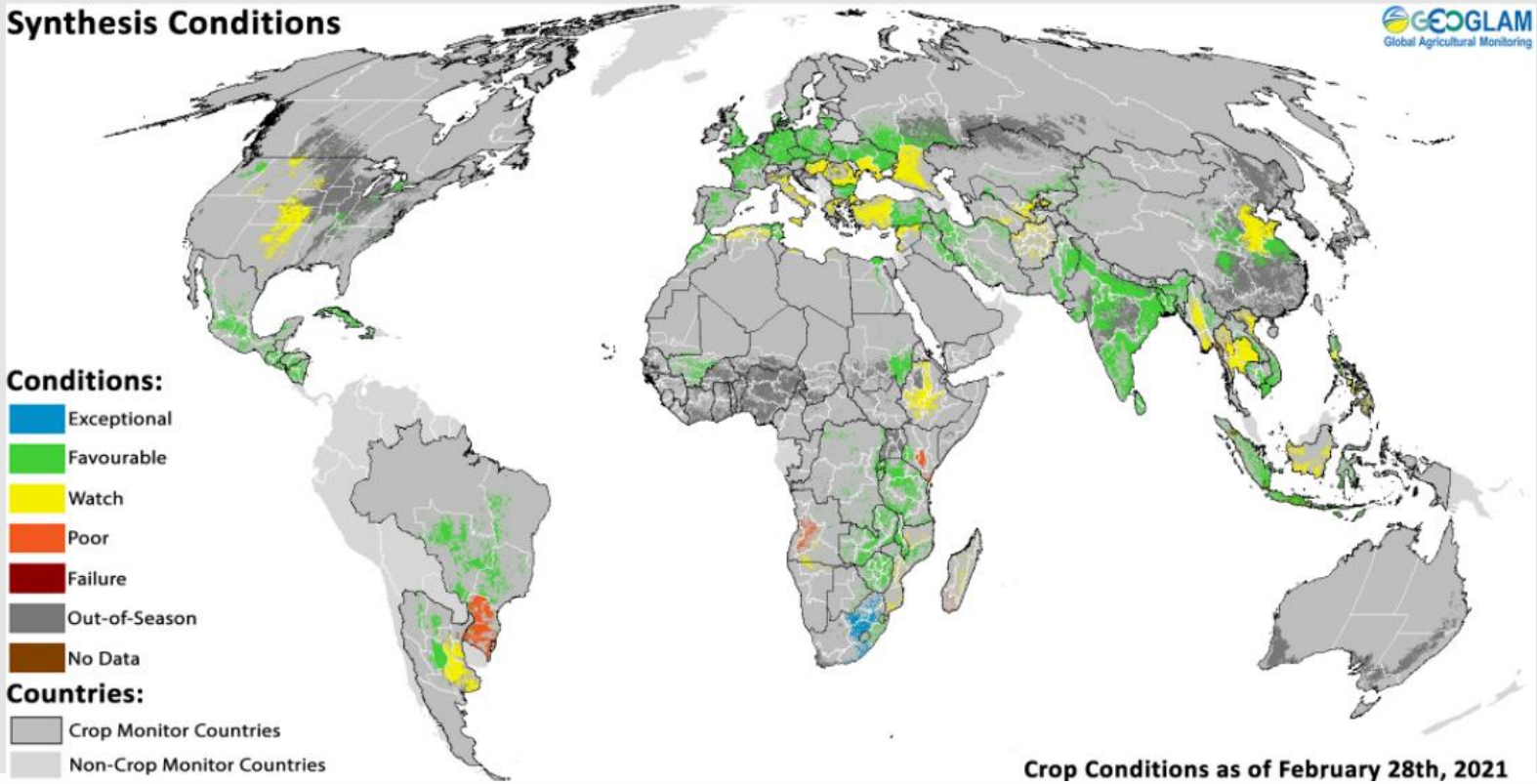


Source: https://www.imf.org/en/Publications/WP/Issues/2019/04/09/Illuminating-Economic-Growth-46670, https://www.imf.org/en/Publications/fandd/issues/2019/09/satellite-images-at-night-and-economic-growth-yao

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/

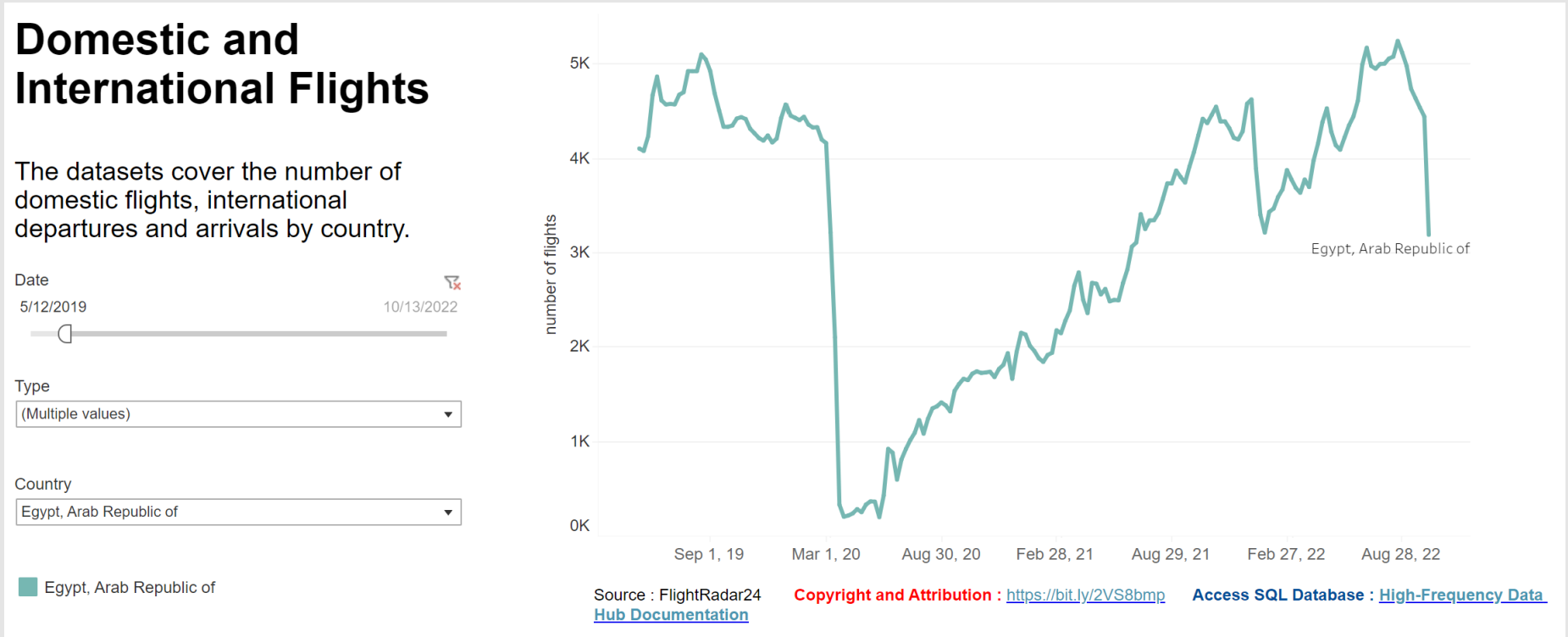
Flight data - tourism activities estimate

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



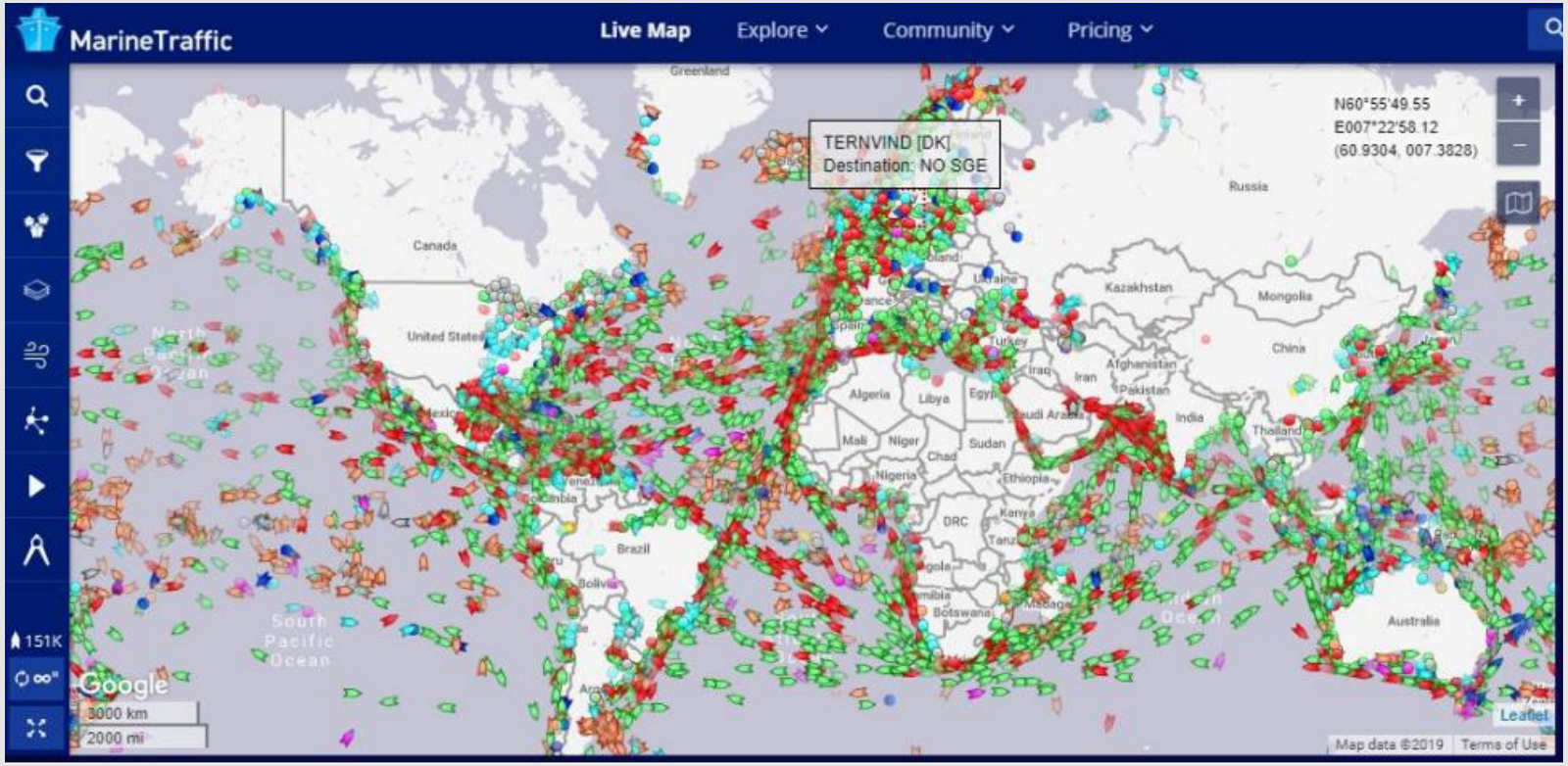
Source: IMF Staff estimates based on Flightradar24 data
Note: Last updated on
https://blogs.imf.org/2020/05/26/keeping-economic-data-flowing-during-covid-19/

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

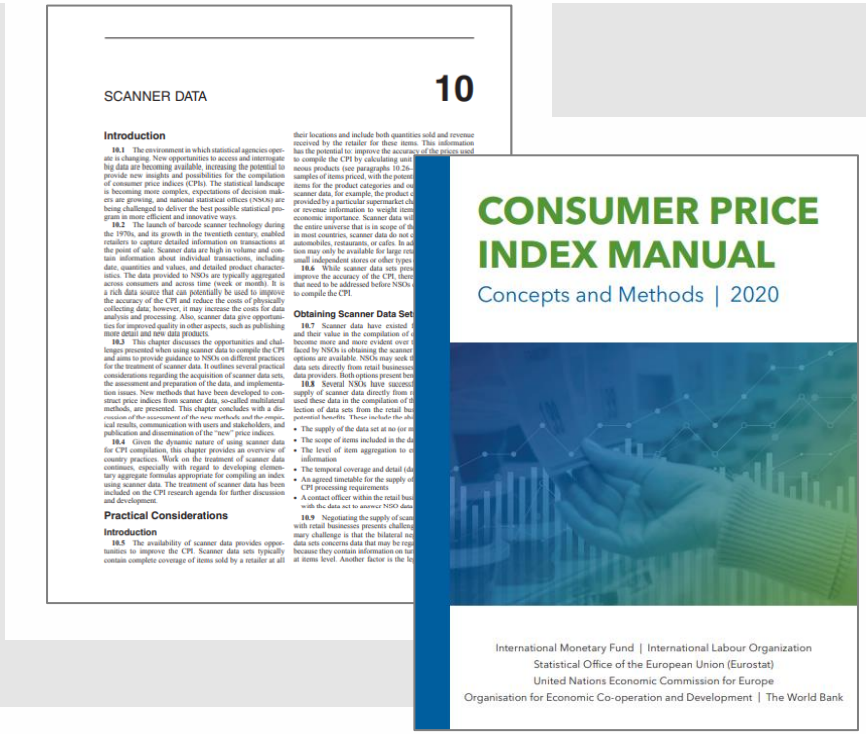


Source: MarineTraffic.
Note: Different types of vessels are shown in different colors.

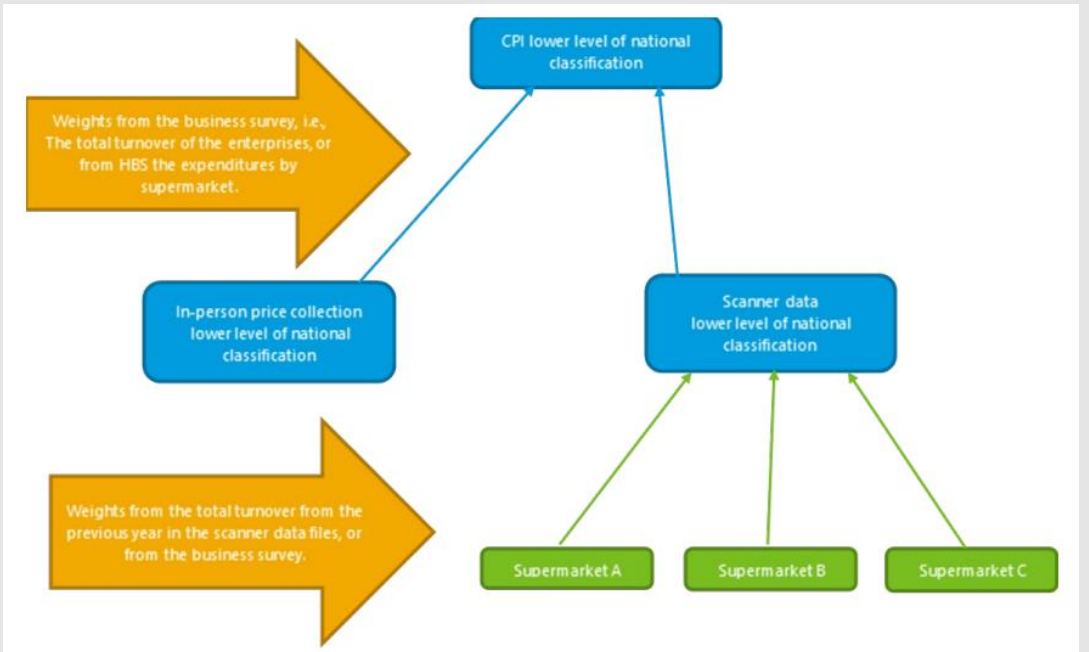
Scanner data for price statistics compilation

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



Source: https://0-www-elibrary-imf-org.library.svsu.edu/view/journals/002/2022/141/article-A001-en.xml?ArticleTabs=abstract

Web scraping

- Examples:
  - Consumer price index (CPI) e.g., hotel reservations, clothing
  - Residential property price index (RPPI) – real estate listings
  - Producer price index (PPI) e.g., product characteristics
  - Vacancies for job statistics
  - Social mood on economy / consumer confidence
  - To improve business registers with more metadata

Case study – using web scraping for the RPPI

Advantages

- Data can be obtained easily
- Geographic coverage
  - Typically most of the country is covered by sites
  - Excellent coverage of important cities
- Characteristics tend to readily available
  - Goal is to give potential buyers as much information as possible
- Timeliness
- Cost efficient compared to traditional survey
- Reduces administrative burden

Case study – using web scraping for the RPPI

Challenges

- Conceptually it is not a transaction price
  - Normally transaction price will be lower than listed price
- Delisted dwellings ≠ sold
- Ghost listings
- Might not be exhaustive
  - Lower-end dwellings might no be found online
- If a website changes drastically your code needs to be adapted

Conclusion: a good starting point while the quality of administrative data improves

Benefits of using new data sources

- **Nowcast** of traditional economic indicators
- **New indicators** (e.g., RPPI)
- **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