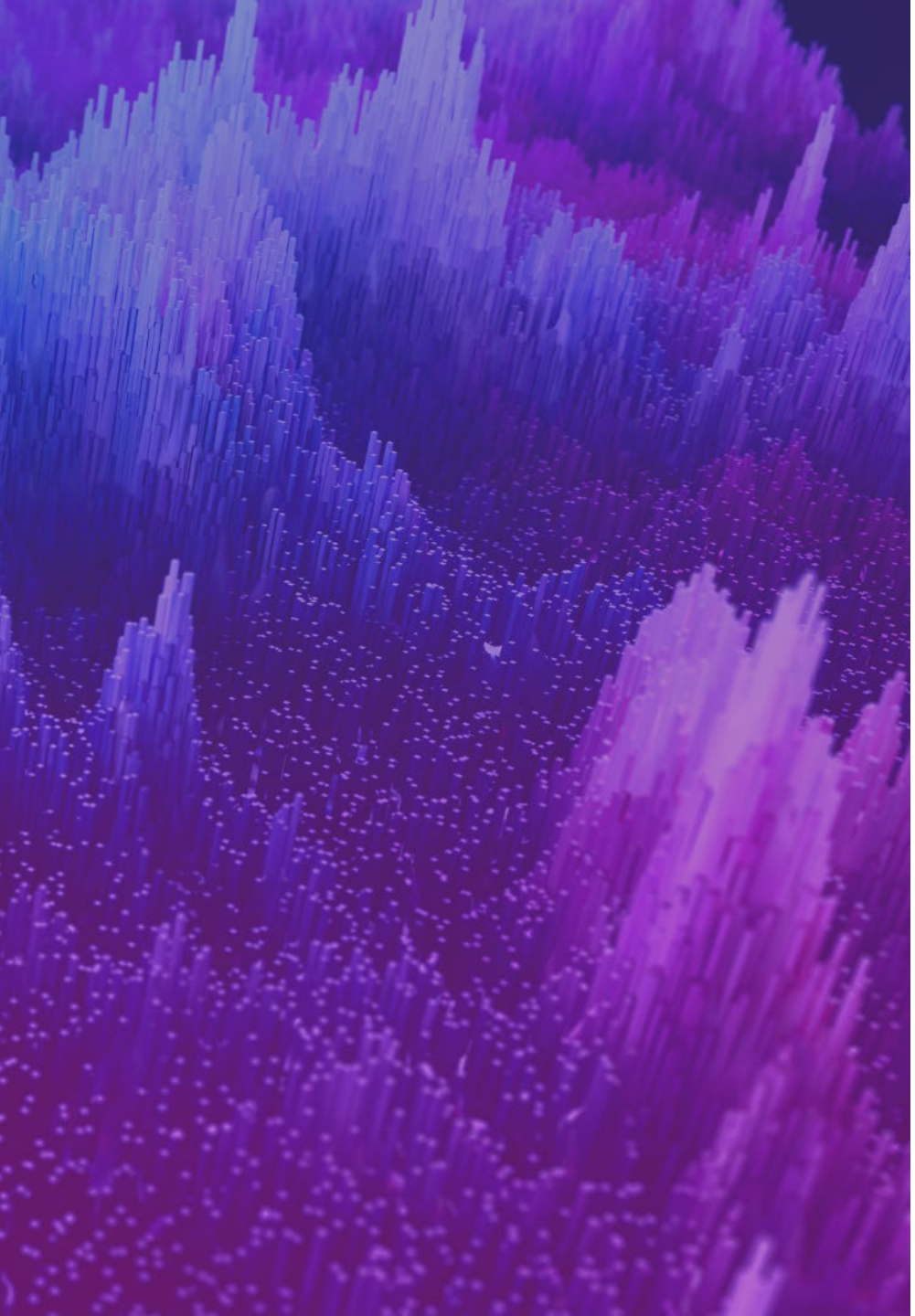


STATISTICS

Use of Big Data in Producing Statistics

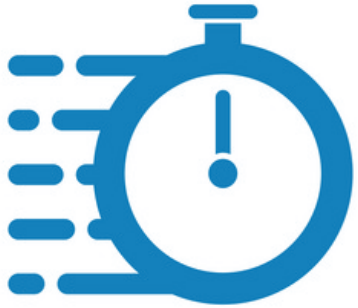
JUNE 19, 2024
D4D FUND STEERING COMMITTEE MEETING

Mr. Seung Mo Choi
Deputy Division Chief, Data Governance and Services Division
IMF Statistics Department

What is Big Data?

- **Big Data** encompasses new data sources, usually by-products, as opposed to survey-based data sources which are designed for production of official statistics.

Very large (volume)	**Quickly produced** (velocity)	**Different in types** (variety)



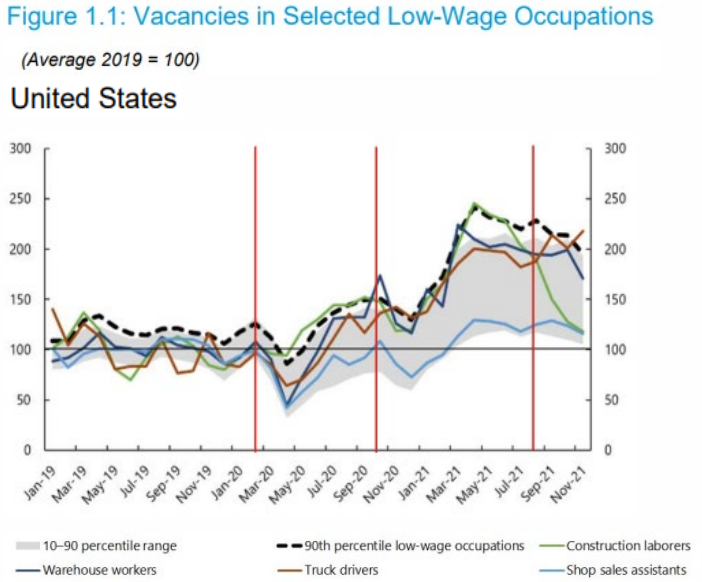
- **Examples of Big Data:** Satellite images, vessel traffic, postings in social networks, search volumes, …
- Broadly, nontraditional data sources include **administrative data** (value added tax, social security, personal and corporate income tax, customs…)

Advantages of Big Data for Macroeconomic Statistics

Timeliness

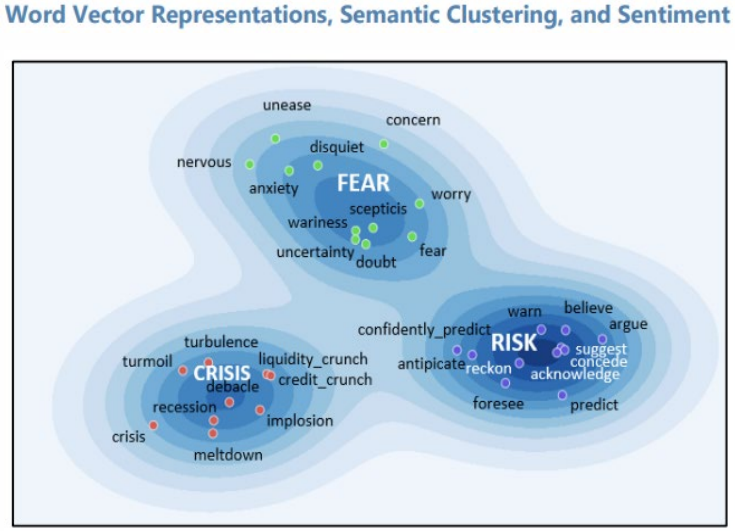
Source: Choi and Iyer (2022).

Granularity of the analysis



Source: Duval and others (2022).

Filling the areas where official statistics are lacking



Source: Huang and others (2019).

Big Data Creates Leapfrog Opportunity in Macroeconomic Statistics

- More countries adapt **technologies** to their statistical compilation programs, with a growing set of:

Digitized administrative data

Primary data from the private sector

- Transforming such data into primary data to produce macroeconomic statistics (as well as policy-relevant climate indicators) requires a specific set of technologies and skills.
  - Traditional surveys are costly and response rates are falling.
  - Demand for CD to explore use of Big Data for official statistics in LLMICs is expected to grow.

A Big Data Center to Support IMF Membership Using Big Data Sources and Analytics

Support responsible and impactful use of Big Data

Positioned as an internationally renowned Centre

Support statistical innovation for IMF and its membership

Role of D4D-II in the use of Big Data

D4D-II would assist countries in integrating Big Data into official statistics programs. The integration of **Big Data** sources into tailored CD will support both the **macroeconomic** and **environmental and climate change** statistics workstreams**.**

MODULE 1
Addressing Data Needs to Empower Policymaking

- The better use of and increased access to administrative data and Big Data to address source data gaps.
- Support countries in processing available Big Data sources to produce high-frequency indicators.

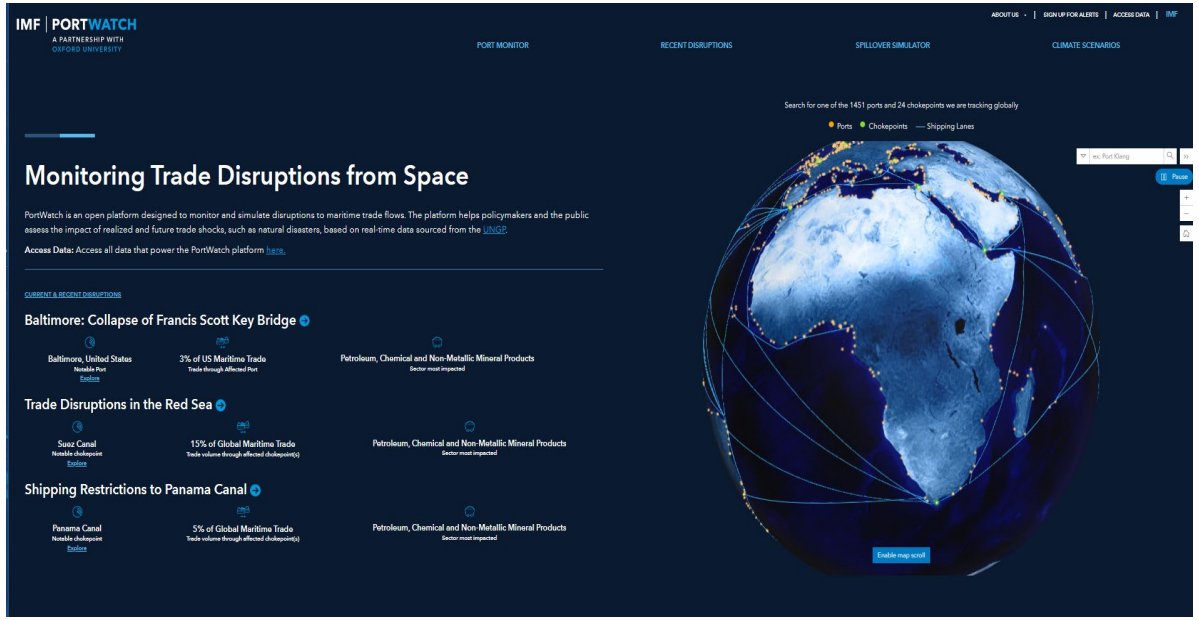
MODULE 5
Environmental and Climate Change Statistics

- Help LLMICs to use Big Data including to obtain estimates of relevant macro critical indicators.
- Big Data sources will also be harnessed to compile climate indicators for several purposes including to support the IMF Resilience and Sustainability Trust.

Role of D4D-II in the use of Big Data: Examples

MODULE 1

Automatic Identification System (AIS) Data and Trade Nowcasting



- ***AIS Data*** (on vessel traffic) have been used in tracking international trade in real time.
  - PortWatch aims to release AIS-derived **real-time trade estimates** to monitor disruptions from shocks, assessing their regional and global spillovers through port-to-port linkages.

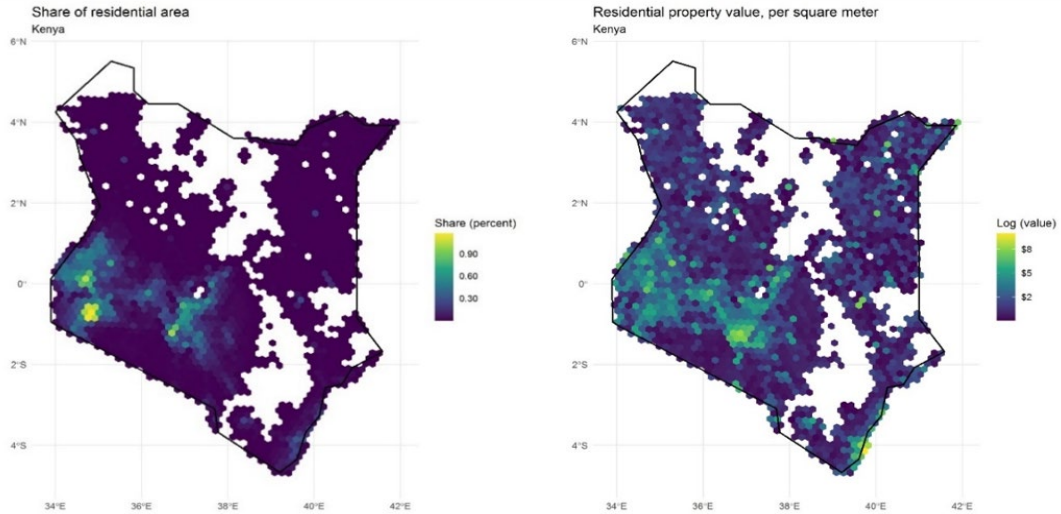
Role of D4D-II in the use of Big Data: Examples

MODULE 5

Physical risk indicators and global data sets

- Physical risk indicators combine information on **hazards, exposures**, and **vulnerability** to help policymakers better understand the risk climate hazards pose to populations and the economy.

- Global data sets can be combined with national and regional data, on the specific vulnerabilities to support designing appropriate adaptation policy and allocating investments in resilient infrastructure where needed.



IMF is piloting a new workshop on “Big Data for Macroeconomic Statistics” for African countries

Jointly organized by AfDB Data Innovation Lab and the IMF Big Data Center.

September 23-27, 2024, in Kigali, Rwanda.

30 representatives of statistics agencies and central banks from **14** African countries will be invited to attend the workshop.

- Contents will include:
  - Key technologies and advanced geospatial analysis techniques and methodologies.
  - AIS data for trade nowcasting (based on the IMF PortWatch platform).
  - NO2, nightlights, vegetation indices, and Google Trends for GDP nowcasting.
  - Advanced natural language processing tools and machine learning models to analyze text data from social media, forums, and customer feedback.

Participants will have the opportunity to present and share knowledge about their use cases with Big Data sources.

Thank you!

Questions?