DATA AT THE WORLD BANK

MOVING BEYOND THE HYPE

WORLD BANK GROUP

Pinelopi Koujianou Goldberg
Chief Economist
World Bank Group

February 13, 2019

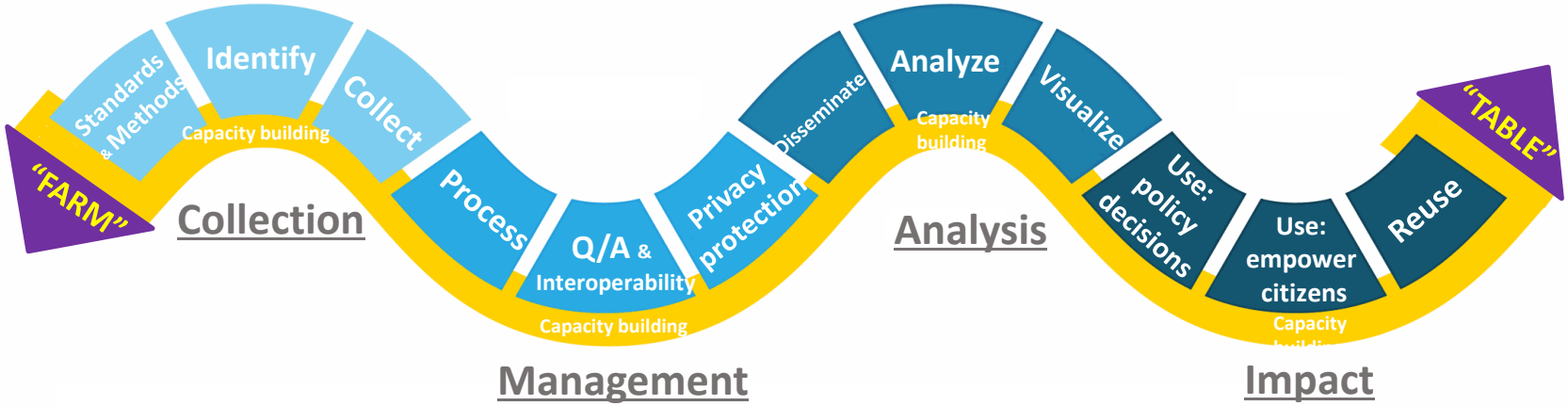
DATA AND THE BANK

The Bank → A leader in data collection, management, and analysis

WORLD BANK GROUP

STEPS IN DATA VALUE CHAIN

Collection → Processing, Storage, Access → Analysis



Based on Open Data Watch schematic prepared for Data2X

WORLD BANK GROUP

DATA COLLECTION: PHILOSOPHY

Two Approaches:

1. Questions → Data? → Go get it
2. Data → Questions?

EXAMPLES OF QUESTIONS → DATA (TRADITIONAL APPROACH)

- ICP
- Doing Business
- Human Capital Project

Main advantages of WB:

- Ability to negotiate with big players, especially governments.
- Scale to harmonize and normalize data.

WORLD BANK GROUP

DATA → QUESTIONS
(BIG DATA)

- The **3** V's of Big Data:
  - Volume
  - Velocity
  - Variety
- Ideally **2** more V's:
  - Veracity
  - Value

WORLD BANK GROUP

BIG DATA (contd.)

- With big data, machine learning techniques become important.
- Three broad categories:
  - Data in public domain (satellite imagery, web).
  - Administrative data: Government is the gatekeeper.
  - Private sector data: Companies are the gatekeeper.
- Role of WB:
  - Use the data in the public domain.
  - Use its soft power to negotiate access to administrative and private data.

SKEPTICISM FROM TRADITIONALISTS

- Too much hype.
- Focus shifts from questions/issues to techniques and algorithms.

MAIN QUESTION

What can we learn from these data that we cannot learn using traditional data?

A POSSIBLE TAXONOMY

- Imagery (satellite, photos, videos).
- Geo-locational data (*e.g.* CDR, Uber/Lyft).
- Network data (*e.g.* Linkedin, Facebook).
- Transactions and Price Data (*e.g.* credit cards, Amazon, Alipay, BPP).
- Text mining (*e.g.* news, Google searches, tweets, text messages, e-mails).

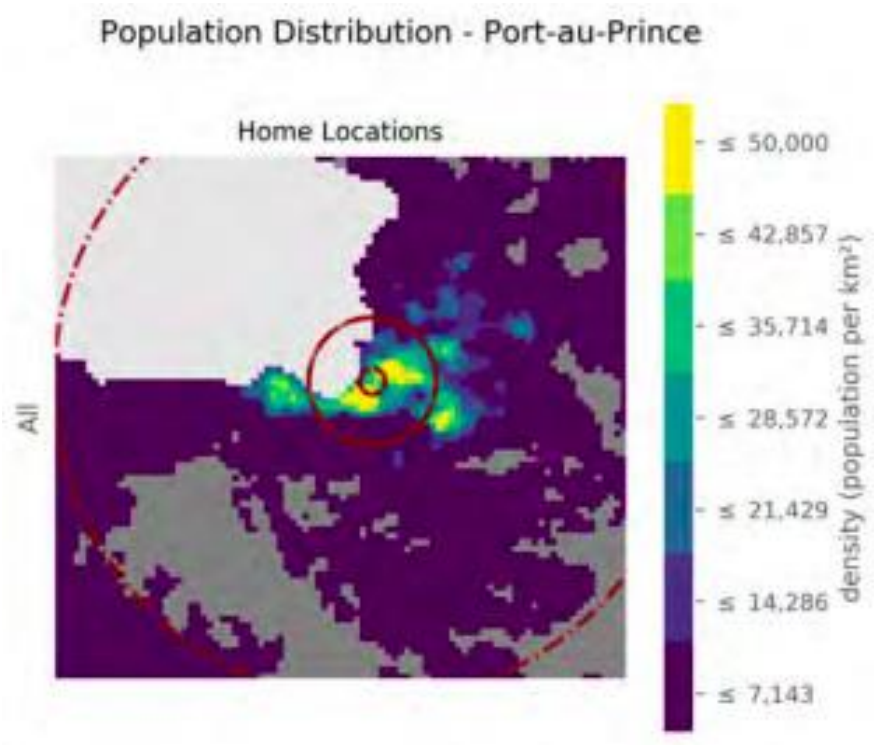
WINS SO FAR

- Real-Time Census and Statistics:
  - Count lights, structures, roads, crops, solar panels (satellite, geo-locational).
  - Estimate inflation (scrape the web for prices → BPP).
  - Predict economic events, *e.g.* recession (text mining).
- Especially valuable in settings where:
  - No information.
  - Inaccurate information.
  - Information purposefully manipulated (*e.g.* BPP's estimate of inflation in Argentina).

WORLD BANK GROUP

USE OF CALL DETAIL RECORDS (CDR) IN HAITI URBANIZATION REVIEW

Residential Population in Port-au-Prince

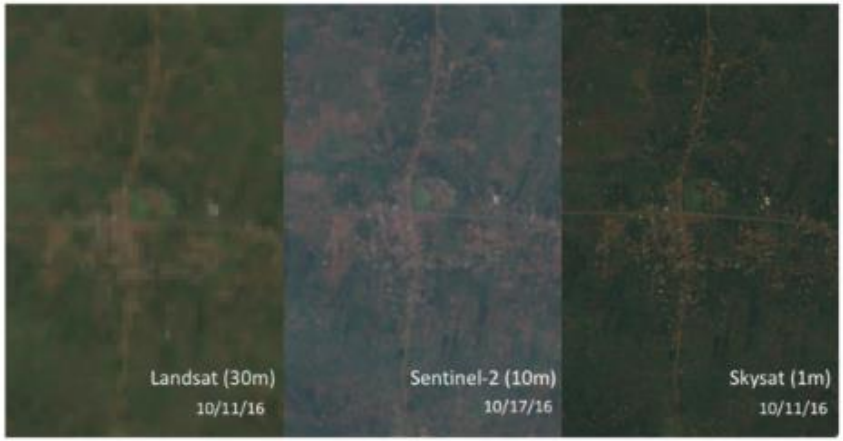


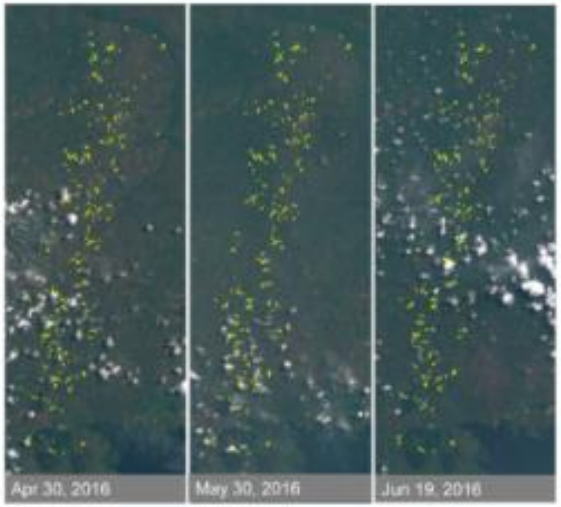
Source: Haitian Cities : Actions for Today with an Eye on Tomorrow, 2017

- **Data:** Call Detail Records (CDRs).
- **Used for:**
  - Identification of key intra-city connectivity challenges.
  - Producing employment accessibility analysis for Port-au-Prince and Cap-Haitien, and;
  - Identifying major bottlenecks and possible interventions, like infrastructure,-re-zoning.

WORLD BANK GROUP

USING SATELLITE IMAGERY/REMOTE SENSING TO PREDICT CROP YIELDS

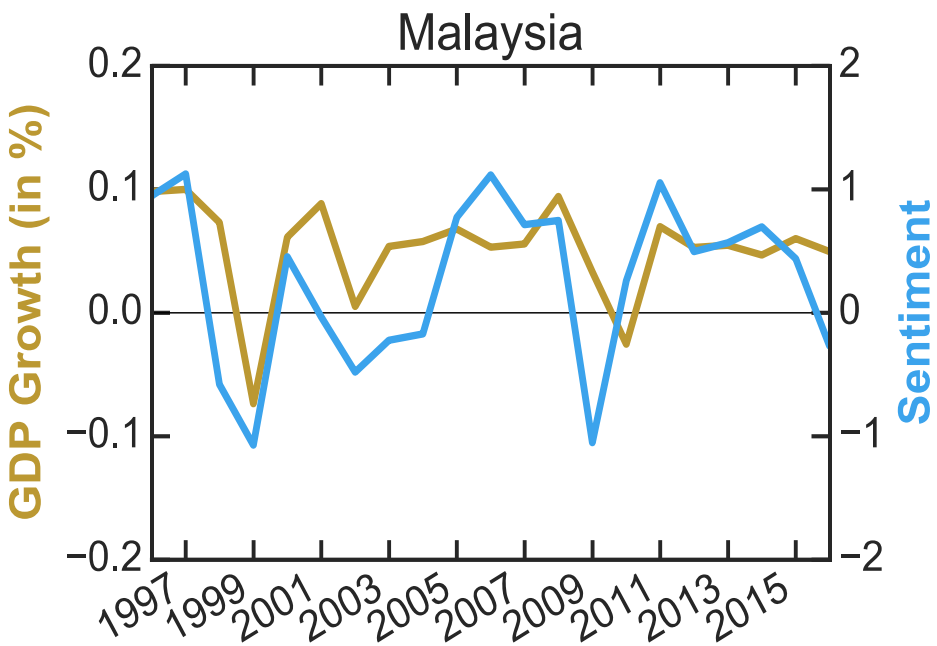




- **Data:** Sentinel-2 imagery-based, remotely-sensed plot-level maize yields with respect to ground-based measures relying on farmer self-reporting in Uganda.
- **Used for:** More accurate, timely and affordable agricultural statistics.
- **Findings:** Methods based on remote sensing could supplement standards approaches for estimating crop productivity.

WORLD BANK GROUP

USING MEDIA ARTICLES TO IMPROVE MACRO FORECASTING



- **Data:** 4M Reuters articles across 25 countries describing economic events over the past 35 years.
- **Used for:** Tracking economic fluctuations, predicting turning points.
- **Findings:** Media sentiment improves predictions of GDP variations by ~15% across countries, compared to consensus forecasts. Used in the Macro Economic Monitor in Malaysia, with plans to scale to many countries.

WORLD BANK GROUP

WINS FOR TOMORROW

Data → Potential Game Changer

- Network and Transactions Data allow unprecedented access to information.

- Digital IDs will allow governments to see what has been invisible so far.
  - ➢ WB could help ensure that this data is used to improve welfare and not to assert government control.

SOME APPLICATIONS…

- Use video data for better monitoring and service delivery.
  - Example: Monitor school teachers to address ‘absentee problem.’
- Use network data to study and affect:
  - Spreading of technology.
  - Employment opportunities.
  - Dissemination of health-care information.
  - Contagion of communicable diseases.
- Use data from digital IDs to identify the poor and better target poverty alleviation.

RISKS AND LIMITATIONS

- Embarrassment of riches. Often data cannot be processed on traditional computers, and statistical and econometric modeling is handicapped by the size of the data.
- Elephant in the room: Privacy.
  - In principle, we can observe everything about a person's life. Data can be used to help or to hurt people.
  - People less sensitive to this risk in developing countries. Price worth paying in order to reduce poverty.
  - But particularly relevant when policy makers are corrupt.
- Bank could play important role in ensuring that data does not get misused.

PROCESSING, STORAGE, ACCESS

- Centralize and coordinate data functions across the Bank.
  - Data Council → A step in the right direction.
- Documentation and Replicability.
  - Make source micro data and code publicly available.
- Access.
- Big issue: Privacy and confidentiality.

WORLD BANK GROUP

DATA ANALYSIS

- Prevalent in every part of the Bank.
- Need feedback from analysis to collection.
- Has always been a strength of the Bank.

PRIORITIES FOR THE BANK

- Invest in firm-level data.
- Invest in data documentation, accessibility and replicability.
- To accomplish the above, invest in people.
  - More statisticians, data-experts.
- Stay atop of new technologies and data, but be aware of hype *vs.* substance.
- New data would be most useful if Bank could use its reputation and soft power to guarantee privacy and confidentiality.
- Use the Bank's soft power to engage in productive partnerships with governments and companies to get access to 'big data' while not compromising privacy.

WORLD BANK GROUP

BUT MOST IMPORTANT:

- Do not lose sight of the questions we need to address.
- Ask: How can we use the ‘big data’ to answer these questions?
- If the ‘big data’ alone is not enough, go collect what is needed.

THANK YOU

WORLD BANK GROUP