# SDGs Amenable to Measurement using Big Data: Examples and Challenges

Rohan Samarajiva & Thavisha Perera-Gomez
2017 International Conference on SDG Statistics
Manila, Philippines
October 5, 2017

LIRNEasia
Pro-poor. Pro-market.

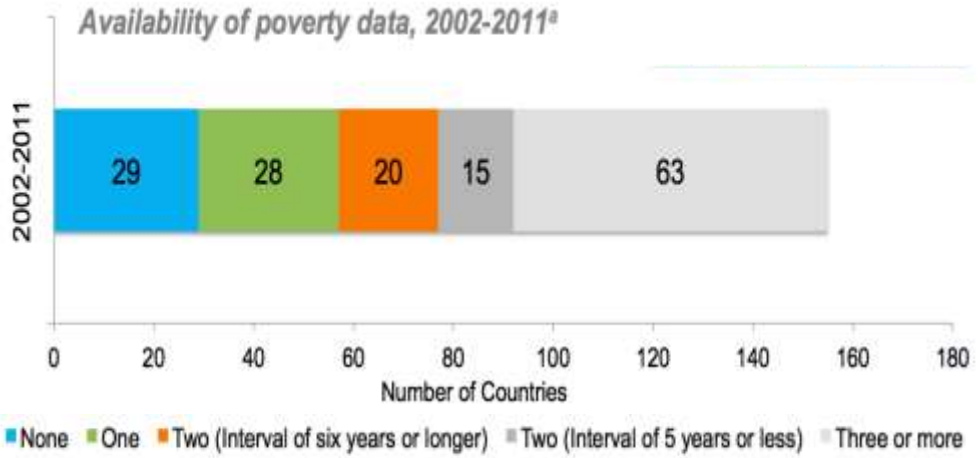
IDRC CRDI
Canada

This work was carried out with the aid of a grant from the International Development Research Centre, Canada

# Unprecedented data requirement to measure progress towards achieving the goals

# Data requirements illustrated

- **How do we know progress is being made toward Goal 1, No Poverty?**
  - Timely, accurate, disaggregate data




- **Call to harness a data revolution for sustainable development:**
- "An explosion in the volume of data, the speed with which data are produced, the number of producers of data, the dissemination of data, and the range of things on which there is data, coming from new technologies such as mobile phones and the "internet of things", and **from other sources, such as qualitative data, citizen-generated data and perceptions data; A growing demand for data from all parts of society**." *(A World that Counts, 2014)*

*There is opportunity to leverage new data sources, including big data to support the measurement/achievement of relevant goals and their associated targets and indicators*

a. Based on the 155 countries that the World Bank monitors poverty estimates for.
Source: Serajuddin, U., Uematsu, H., Wieser C. Yoshida, N. & Dabalen, A. (2015). Data Deprivation: another deprivation to end

LIRNEasia
Pro-poor Pro-market

# Can we engage in SDG work using datasets that do not include the poor?

- Administrative data
  - E.g., digitized medical records, insurance records, tax records
- Commercial transactions (transaction-generated data)
  - E.g., Stock exchange data, bank transactions, credit card records, supermarket transactions connected by loyalty card number
- Online activities/ social media
  - E.g., online search activity, online page views, blogs/ FB/ twitter posts
- Sensors and tracking devices
  - E.g., road and traffic sensors, climate sensors, equipment & infrastructure sensors, mobile phones communicating with base stations, satellite/ GPS devices

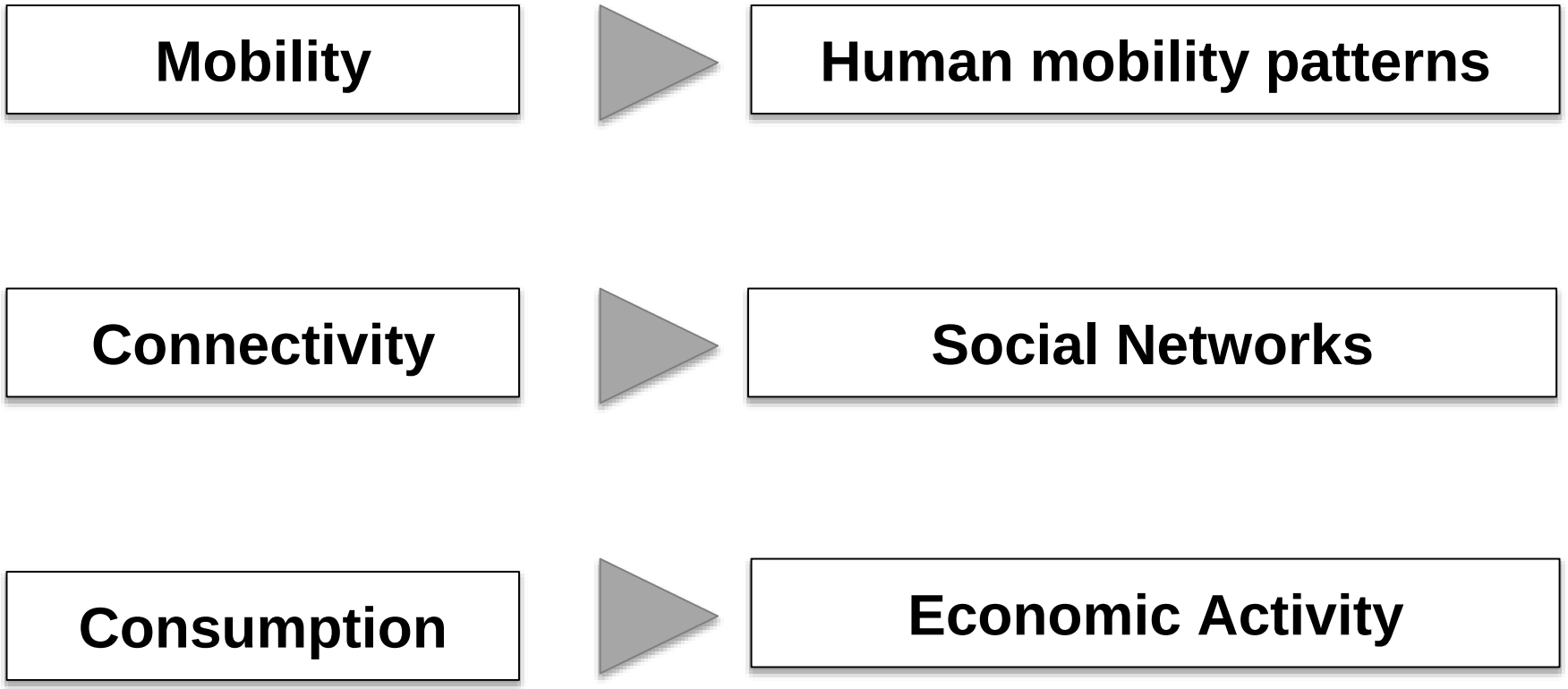
# Mobile Network Big Data is only option for some problems at this time

| Country | Mobile Subscriptions/100 | Internet Users/100 | Facebook Users/100 |
|---|---|---|---|
| | 2016 | 2016 | 2017 |
| Pakistan | 71.4 | 15.5 | 15.8 |
| Bangladesh | 77.9 | 18.3 | 15.8 |
| India | 87.0 | 29.6 | 15.9 |
| Myanmar | 89.3 | 25.1 | 29.2 |
| **Philippines** | **109.2** | **55.5** | **59.7** |
| Sri Lanka | 118.5 | 32.1 | 25.0 |
| Indonesia | 149.1 | 25.4 | 44.4 |
| Thailand | 172.7 | 47.5 | 70.3 |

Sources: http://www.itu.int/net4/itu-d/icteye/AdvancedDataSearch.aspx; http://datatopics.worldbank.org/hnp/popestimates; facebook advertising portal;

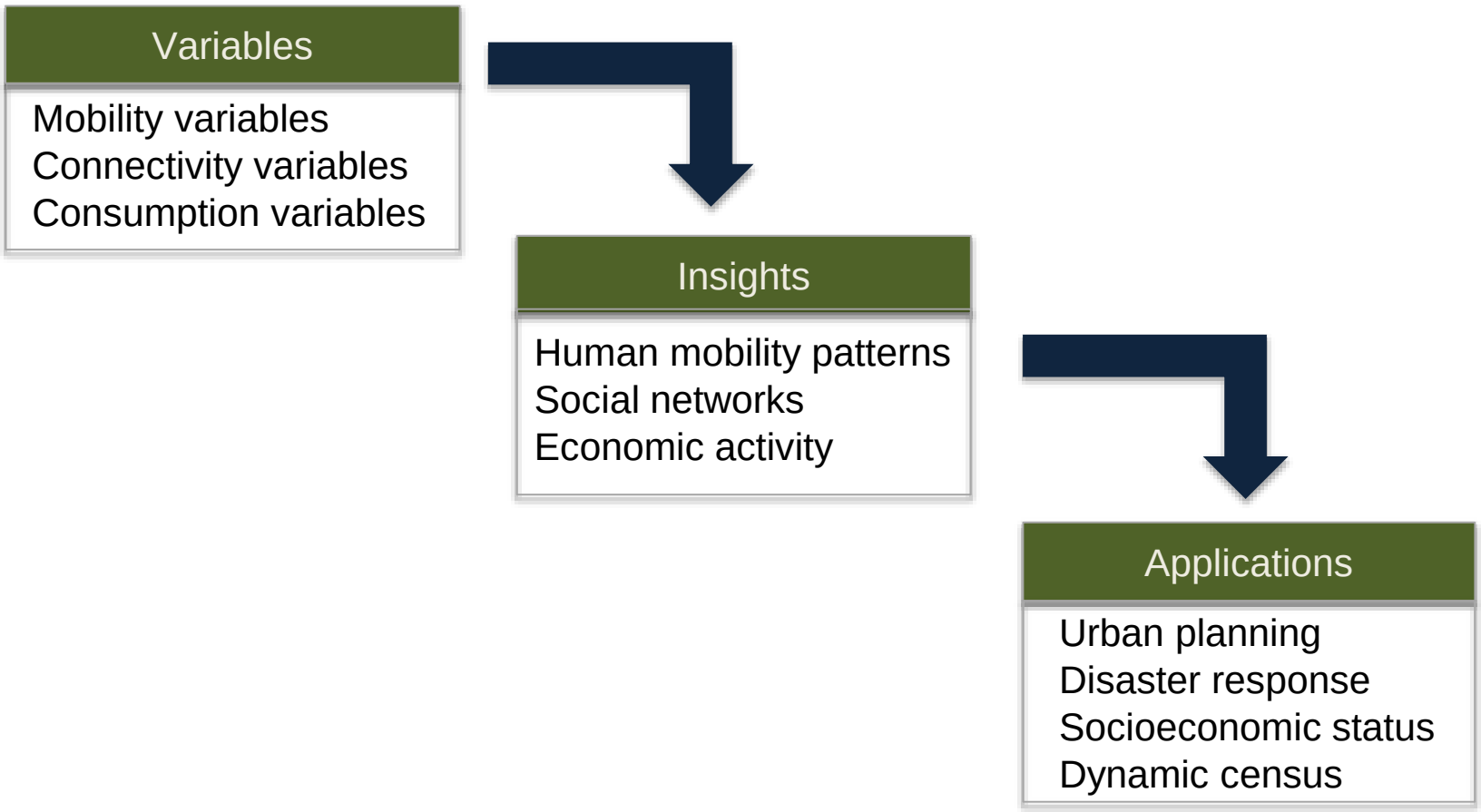
LIRNEasia
Pro-poor Pro-market

# Variables derived from Mobile Network Big Data can used for development purposes



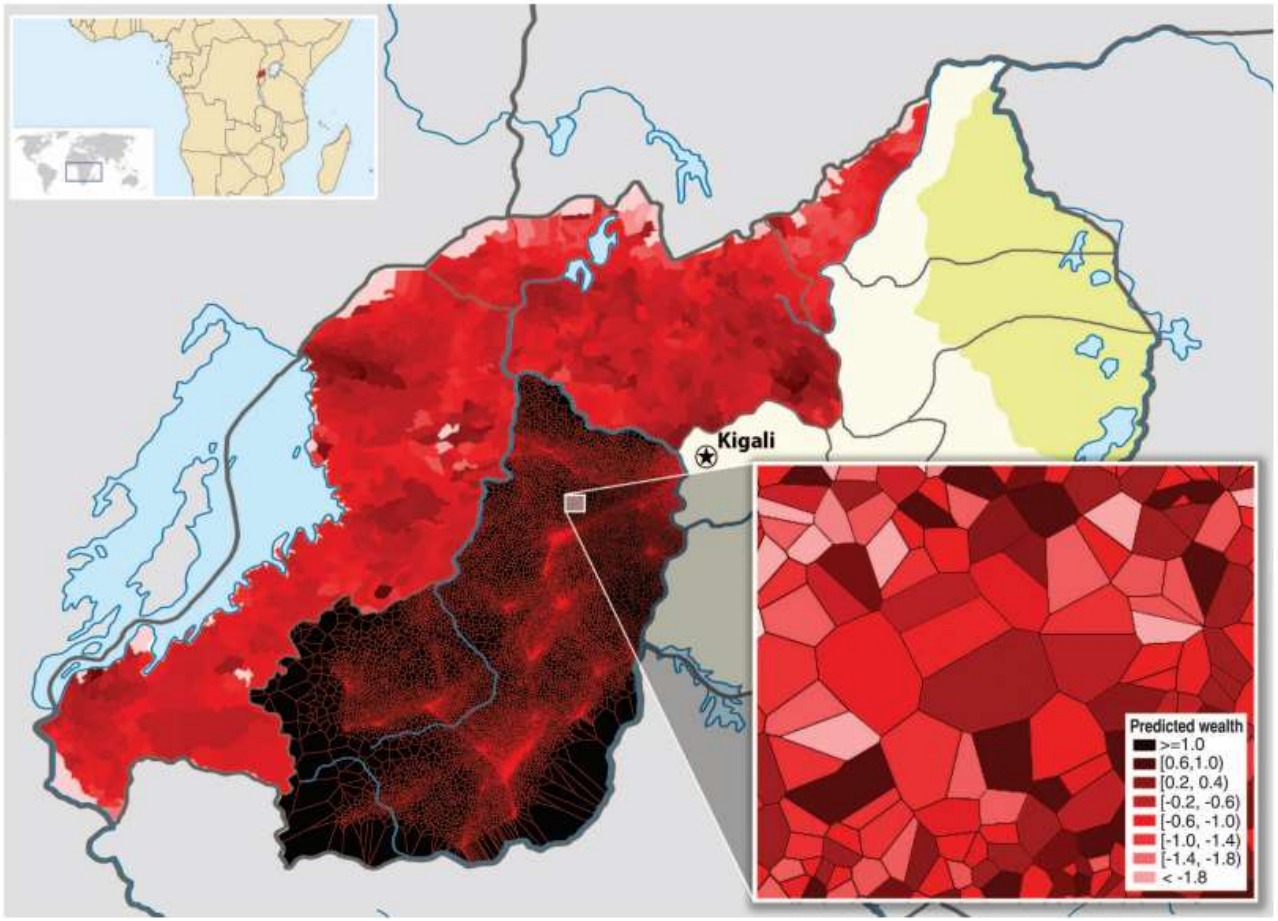
# Insights derived from Big Data can potentially be applied to SDGs

For instance, consider mobile network big data



## ACHIEVE TARGETS

**Target 1.2.** By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions




## IDENTIFYING THE POOR AT MORE GRANULAR LEVELS

Estimating wealth using mobile phone data + Survey data

Predict wealth at a district level, validated with government data

Shows opportunity to predict wealth at micro regions

*Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. Science, 350(6264), 1073-1076. Chicago*

Image source

## ACHIEVE TARGETS

Target 11.2 By 2030, provide access to safe, affordable, accessible and sustainable transport systems for all, improving road safety, notably by expanding public transport, with special attention to the needs of those in vulnerable situations, women, children, persons with disabilities and older persons

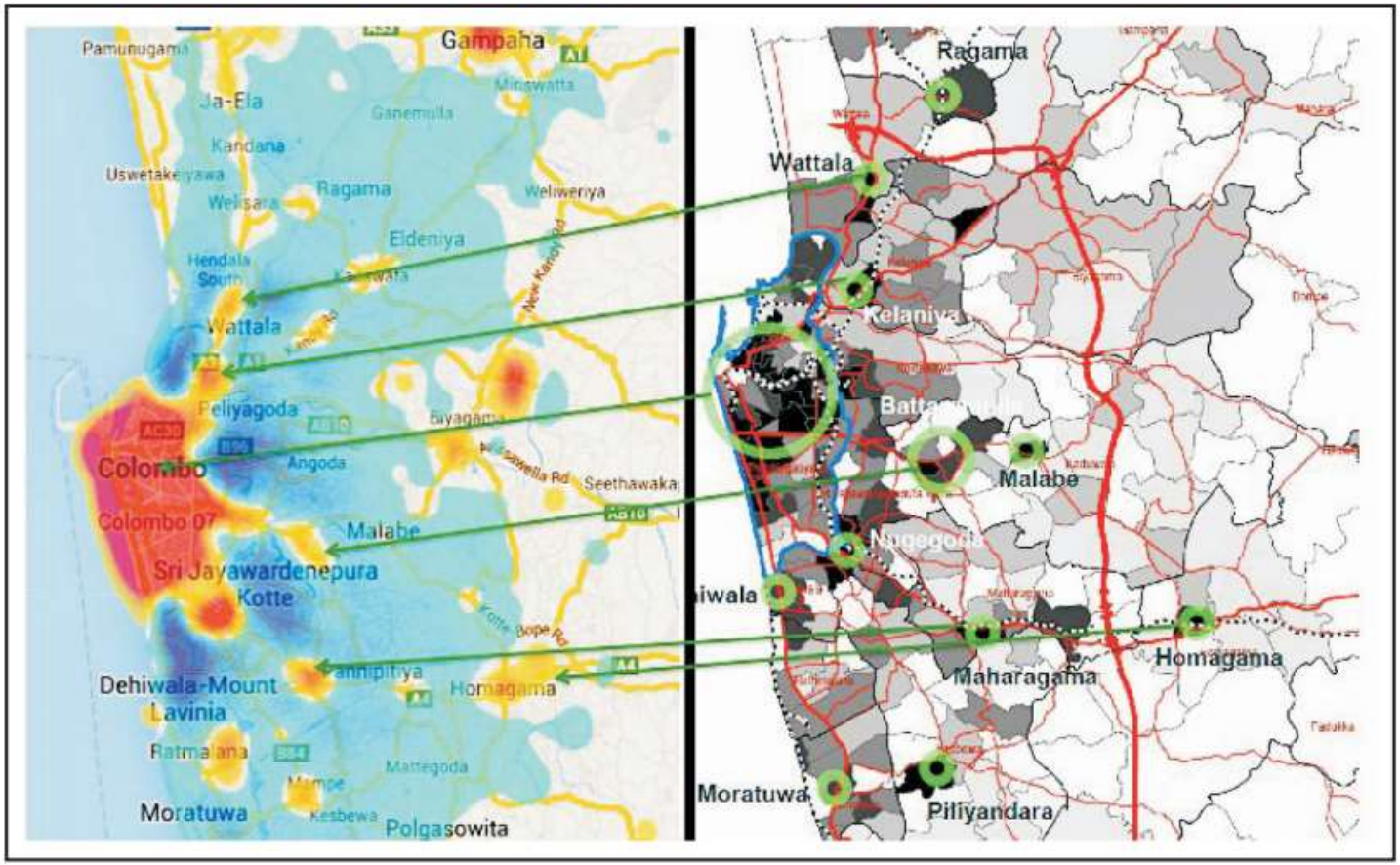



*Figure 5a. Result of analysis of mobile network*

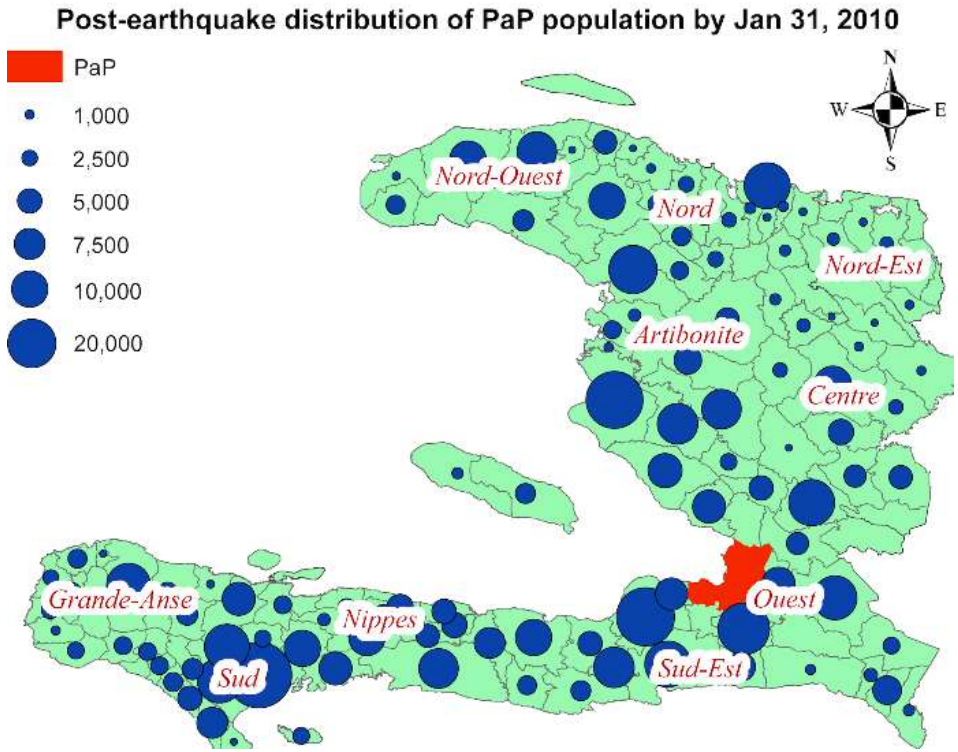
*Figure 5b. Result of analysis of survey data.*

*Lokanathan, S., Kreindler, G. E., de Silva, N. N., Miyauchi, Y., Dhananjaya, D., & Samarajiva, R. (2016). The Potential of Mobile Network Big Data as a Tool in Colombo's Transportation and Urban Planning. Information Technologies & International Development, 12(2), pp-63. Chicago*

## SUPPORT INDICATORS

Indicator 11.5.1. Number of deaths, missing persons and directly affected persons attributed to disasters per 100,000 population




## DISPLACEMENT OF POPULATION AFTER DISASTERS

Bengston et al. (2011) tracked population movements after an earthquake in Haiti using mobile phone network data

*Bengtsson, L., Lu, X., Thorson, A., Garfield, R., & Von Schreeb, J. (2011). Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. PLoS medicine, 8(8), e1001083.*

# The big data for development space offers many opportunities for strategic partnerships

**Multilateral Organizations**

UNITED NATIONS

GLOBAL PULSE

BigData UN Global Working Group

4th UN Conference on Big Data

Bogota, Colombia 8 - 10 Nov 2017

Big Data Project Inventory

**Non-Governmental Organizations/Non-Profits/Civil Society**

DATA POP ALLIANCE

HARVARD HUMANITARIAN INITIATIVE

mit media lab

ODI

FLOWMINDER.ORG

LIRNEasia www.lirneasia.net

**Academia/Research Labs**

Center for Spatial Information Science & Institute of Industrial Science, the Univ. of Tokyo

Shiabasaki & Sekimoto lab

- Joshua Blumenstock (University of California, Berkeley),
- Vanessa-Frias Martinez (University of Maryland)

**Government**

- National statistical institutions

**Data providers**

telenor

Telefónica

# Challenges of leveraging big data for SDGs

## Data Access

- Most troves of big data of value for SDGs such as mobile phone data and social media data are in the hands of competitive private firms

  **Competition issues**

  – Firms are wary about sensitive information leaking out

  **Privacy and Security**

  – Also concerned about negative PR if controversies arise

  – Worries about data leaving the country

  **Transaction costs**

  – Costs of pseudonymizing data, using own resources

  – Costs of making data consistent and anlyzable

  – Costs of legal compliance and approval

# Challenges of leveraging big data for SDGs

**Skills**

- How to find and hold data scientists & multi-disciplinary teams

**Hardware**

- Not really an issue for batch processing, though costly for real-time

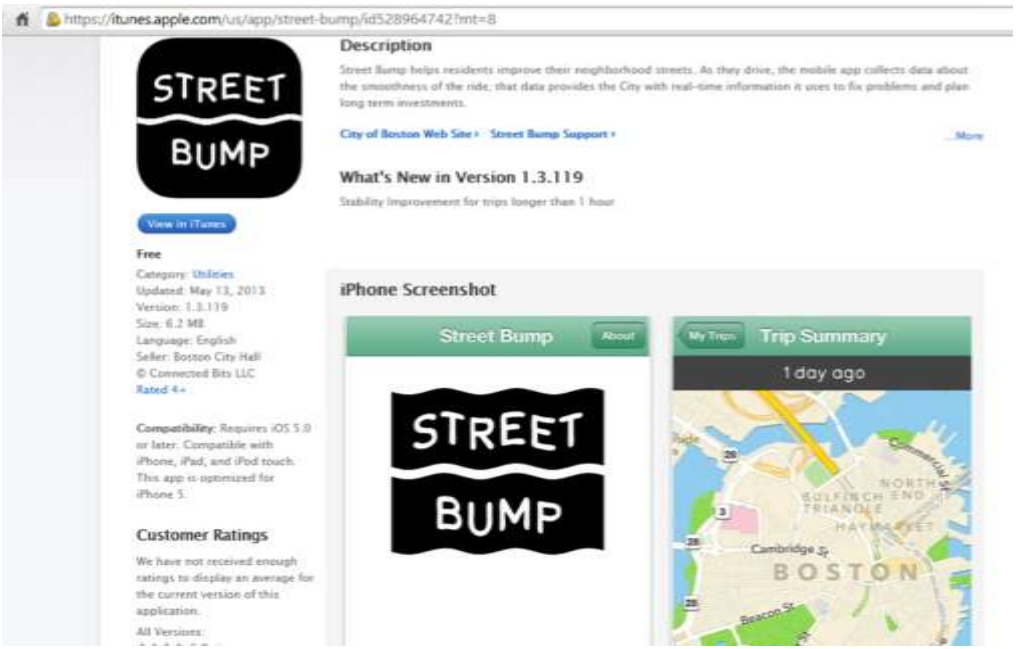
**Software**

- Mostly open source, though rapidly evolving

**Representivity**

- Most important issue for big data used for public purposes

# Bias in big data ➔ why mobile network big data in developing countries




- Streetbump is a Boston crowdsourcing + big data application that uses the natural movement of citizens to improve street maintenance
  - –Data generated from an app downloaded to a smartphone "mounted" in a car
- Can Streetbump be transplanted in Manila at this time?
  - –Feature phones >> Smartphones
- "Something better than nothing" may not apply
  - –Bias toward roads traversed by smartphone owners ➔ In conditions of limited resources, may skew resource allocation