

# The data revolution: Leveraging the rise of big data

Sriganesh Lokanathan

Lessons Learnt Workshop for Countries Designing UN Strategic Development  
Frameworks in 2015 in Asia-Pacific

Bangkok, 25 March 2015



Some of the work mentioned was carried out with the aid of a grant from the International Development Research Centre, Canada and the Department for International Development UK..



# The rationale for the data revolution in development

- Despite recent advances, existing data has problems:
  - The cost means data is not collected frequently enough
  - Collected data is not released fast enough
  - Marginalized populations are often not counted
  - The data is often not sufficient for the needs

**So where do we look for new data to give us better insights, more frequently and at higher resolution?**

**What is big data and where can you find it?**

**What are the benefits?**

**What are the challenges?**

**Who should you be talking to?**

**What is big data and where can you find it?**

**What are the benefits?**

**What are the challenges?**

**Who should you be talking to?**

# What is big data?

- An all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process using traditional data processing applications
  - Qualitatively different from previous forms of data
- Challenges include: analysis, capture, curation, search, sharing, storage, transfer, visualization, and privacy violations
- Examples:
  - 100 million Call Detail Records are generated per day by Sri Lanka mobile operators
  - 45 Terabytes of data from Hubble Telescope

# So where will the data come from?

- 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
- Sensors and tracking devices
  - E.g., road and traffic sensors, climate sensors, equipment & infrastructure sensors, mobile phones communicating with base stations, satellite/ GPS devices
- Online activities/ social media
  - E.g., online search activity, online page views, blogs/ FB/ twitter posts

**We saw many examples earlier of the potential insights from these new data sources**



# However, only mobile network big data currently has broad population coverage in developing economies

	Mobile SIMs/100	Internet users/100	Facebook users/100
Bangladesh	74	7	7
DPR Korea	10		0.03
Iran	84	31	
Lao PDR	68	13	9
Mongolia	124	18	32
Thailand	140	29	42
Vietnam	131	44	27
Sri Lanka	96	22	13

Source: ITU Measuring Information Society 2014; Facebook advantage portal; NSOs

# How do mobile operators see the world?

- **Call Detail Record (CDR)**

- Records of all calls made and received by a person created mainly for the purposes of billing
- Similar records exist for all SMS-es sent and received as well as for all Internet sessions

Calling Party Number	Called Party Number	Caller Cell ID	Call Time	Call Duration
A24BC1571X	B321SG141X	3134	13-04-2013 17:42:14	00:03:35

- The Cell ID in turn has a lat-long position associated with it.

- **Airtime reload records**

- Records of all airtime reloads performed by prepaid SIMs
- Each row corresponds to a record of one person's activity:

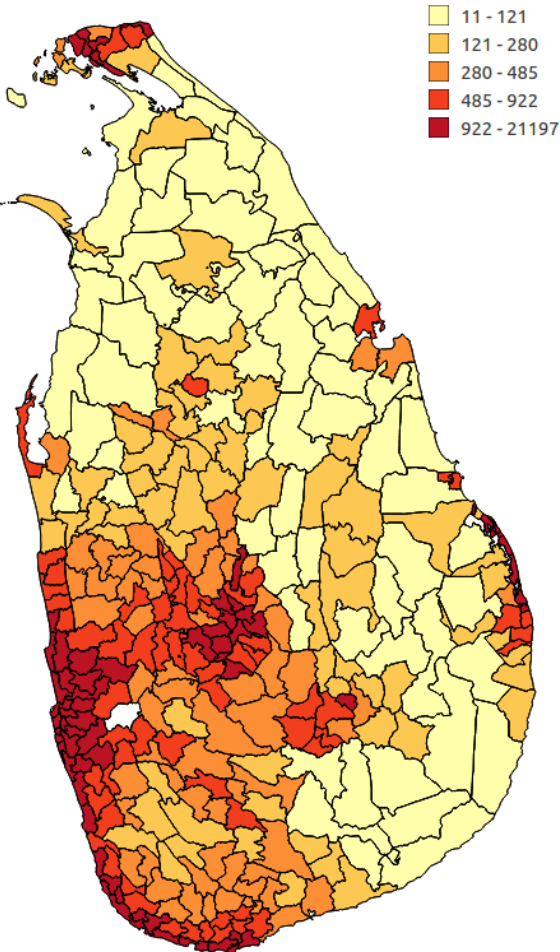
Number	Type of recharge	Starting balance	Amount	Time
A24BC1571X	CARD	0.41	50	13-04-2013 17:42:14



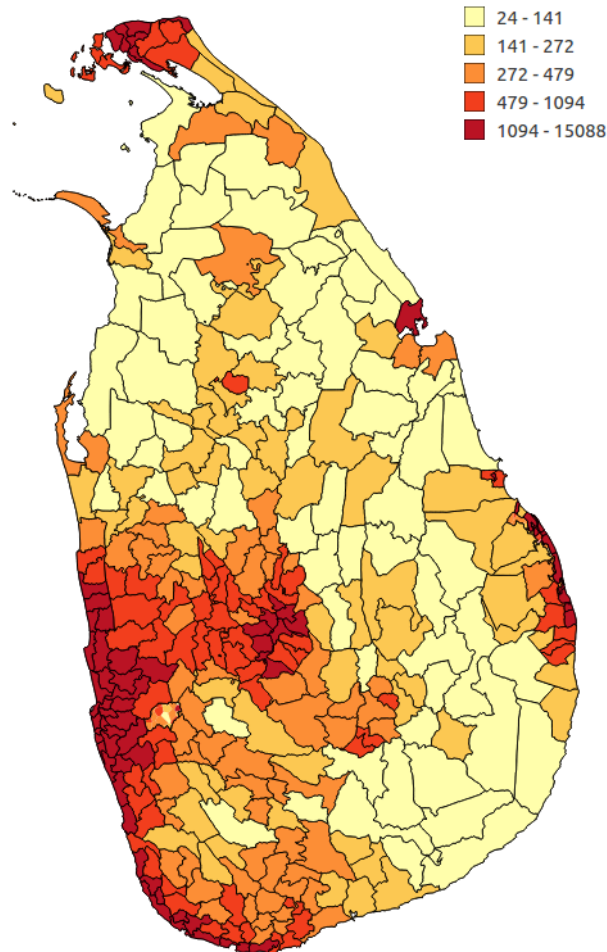
# What can we do with mobile data?

# Produce granular & high-frequency estimates of population density

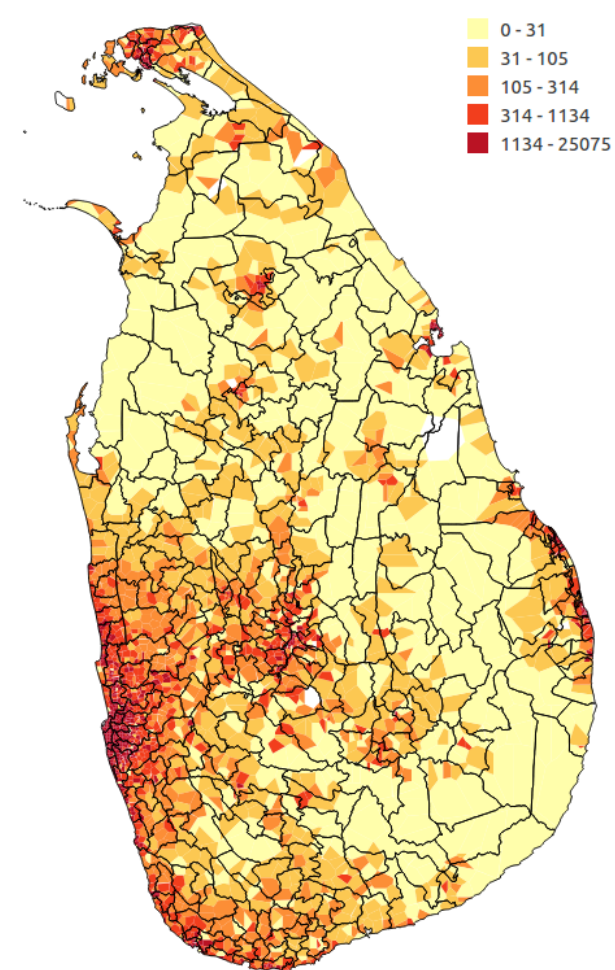
**DSD population density from 2012 census**



**DSD population density estimate from MNBD**

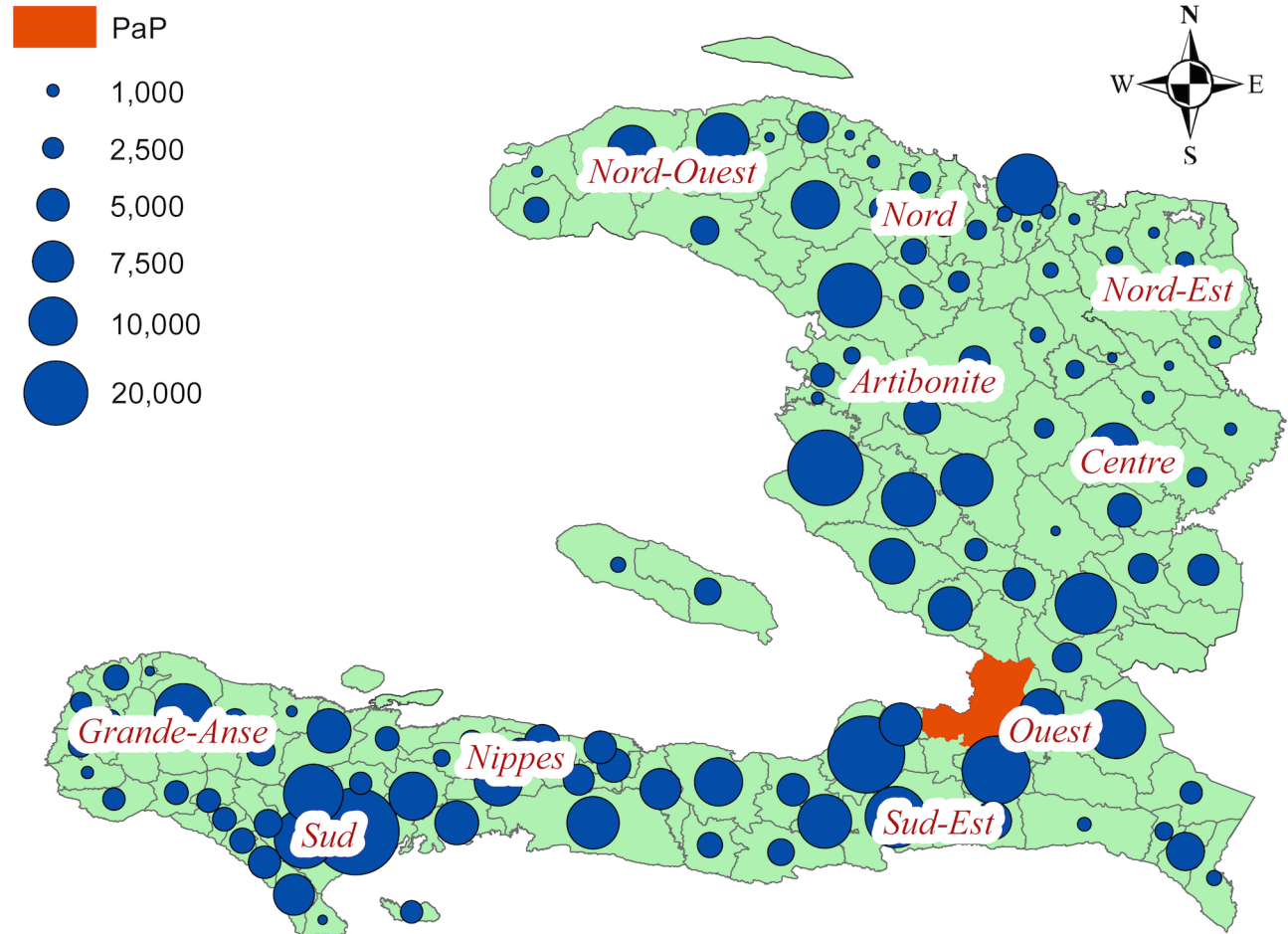


**Voronoi cell population density estimate from MNBD**



# Quickly find displaced populations post disaster

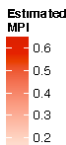
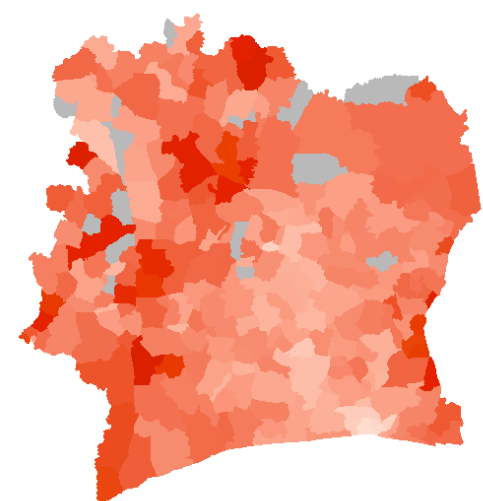
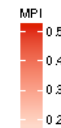
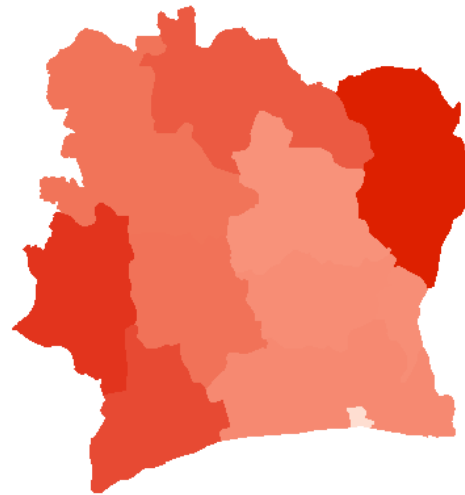
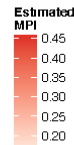
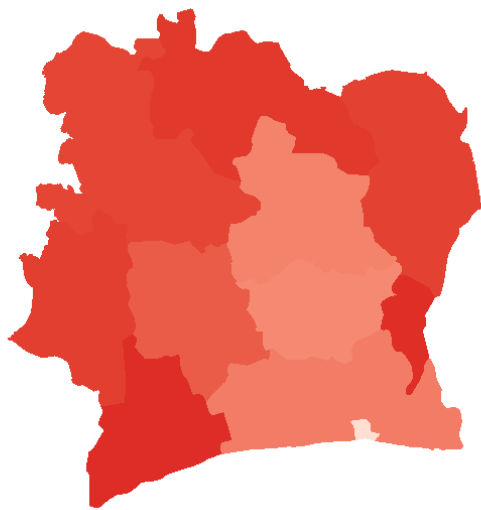
The figure shows the number of people estimated to have been in Port-au-Prince (PaP) on the day of the 2010 Haiti earthquake, but *outside* the capital 19 days later. The circles represent the numbers of people who were displaced. This map was produced on the basis of mobile-network data to show the potential of big data in tracking population movements



**Source:** 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. doi:10.1371/journal.pmed.1001083

# Obtain granular & high-frequency estimates of poverty and socio-economic levels

Lighter color denotes higher poverty



**Multidimensional poverty index (MPI) for 11 regions**

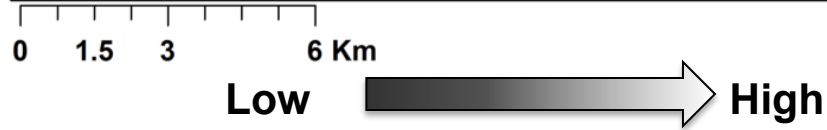
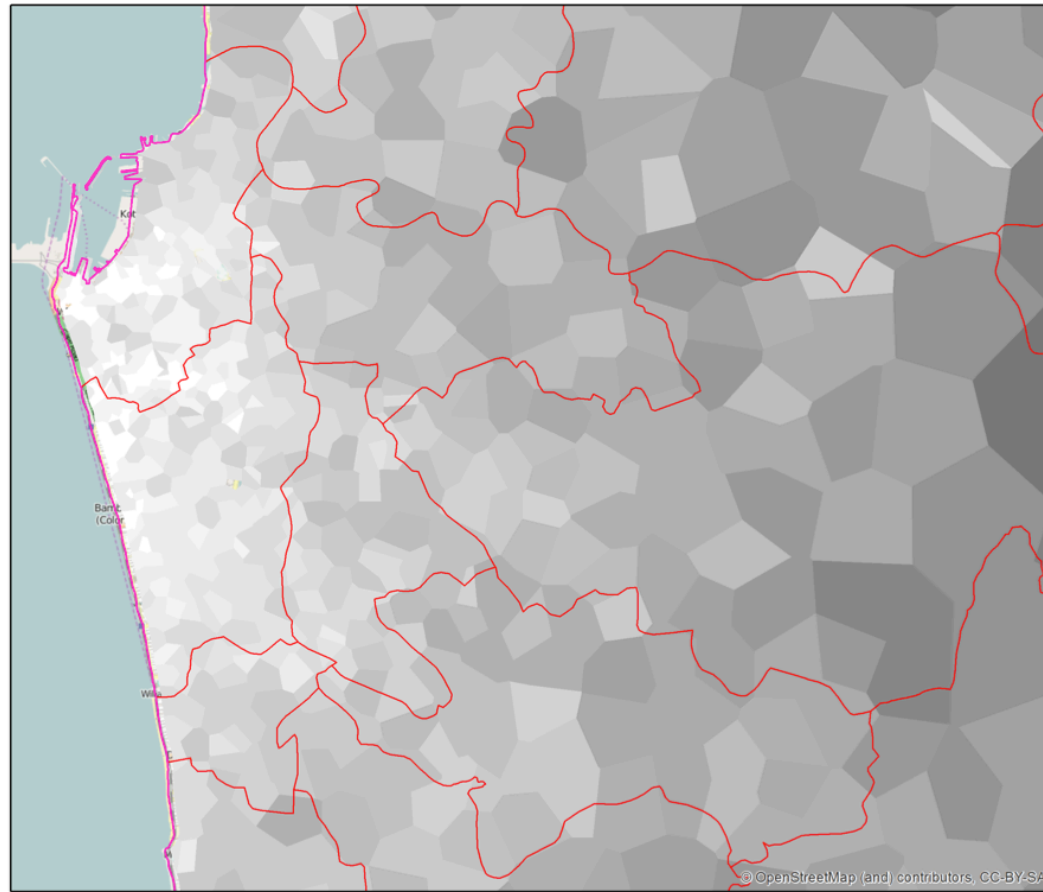
**MPI estimated from diversity of communication diversity between 11 regions**

**Much finer MPI estimation based on tower-to-tower communication diversity**

Source: Smith, C., Mashadi, A., & Capra, L. (2013). Ubiquitous Sensing for Mapping Poverty in Developing Countries Submitted to D4D @ NetMob 2013. In Data for Development: Net Mobi 2013. Cambridge, MA.

# Obtain high frequency proxy measures for economic activity in regions

**Economic activity/km<sup>2</sup>**



## ...and many more

- Understand labor market mobility
- Understand seasonal and long term migration
- Identify displaced communities post disaster
- Model the spread of communicable diseases
- Estimate informal economic activity
- Facilitate financial inclusion for the unbanked through new proxy measures for “creditworthiness”
- Geo-demographics

**These insights will come from ‘meshing’ data including extant official statistics, using new techniques**



What is big data and where can you find it?

**What are the benefits?**

What are the challenges?

Who should you be talking to?

# So what's the real advantage of such new data for the data revolution?

Characteristics	Benefits
High frequency	<ul style="list-style-type: none"><li>• Can complement surveys and censuses with more frequent measurements (e.g. commuting and seasonal migration)</li><li>• “New” findings (e.g. hourly population density, informal economic activity)</li></ul>
High resolution	<ul style="list-style-type: none"><li>• Insights at fine geographic resolution</li></ul>
Near real time	<ul style="list-style-type: none"><li>• Create near real-time proxy indicators (e.g., economic activity, socio-economic levels.)</li><li>• Respond to real-time events</li></ul>
(Almost) universal coverage	<ul style="list-style-type: none"><li>• Allow extrapolation of surveys (e.g. income &amp; expenditure surveys, labor force surveys) to larger samples</li><li>• Capture informal (economic) activities</li></ul>



What is big data and where can you find it?

What are the benefits?

**What are the challenges?**

Who should you be talking to?

# Challenge: accessing government data

- Data from government
  - Requires greater ‘datafication’ especially in developing economies
  - Being spearheaded through the open data movement

# Challenge: accessing private sector data

- Not amenable to open data initiatives
- There can be competitive implications to data release, even if anonymized

## Potential opportunities for accessing data:

- The value of ‘mashing’ data (both private as well as government data) is greater than the sum of the parts
- The state of the art is still developing enabling unique collaborations e.g. UN Global Pulse innovation labs, Orange Data for Development challenges, UP Singapore, LIRNEasia
- There is a potential role for intermediaries that interface between private business interests and public uses e.g. ongoing work at LIRNEasia

# Challenge: new skills

- Statistics can work differently in a big data world
  - When the data is that large, it is easier to find statistically significant predictors even if the effect size is small
- Modeling complex relationships will require machine learning techniques (traditionally from the domain of computer science)

## Opportunity:

- New collaborations
- Inter-disciplinary teams will be essential if we are to leverage such data

# Challenge: protecting privacy

- The difficulty is that:
  - The higher the resolution of your insights, the greater the privacy implications
  - Mixing non-personal data with other sources can reveal personal attributes
  - Informed consent is meaningless in a big data world
  - People often don't know what their generalizable privacy needs are or how their preferences might evolve

## Potential solutions:

- Privacy by design
- Different levels of disaggregation of shared data
- Technical solutions (e.g. anonymization, adding noise)
- Evolution from a rights based approach to a harms based approach

What is big data and where can you find it?

What are the benefits?

What are the challenges?

**Who should you be talking to?**

# What's the role for NSOs?

- Potential roles being talked about:
  - Certify statistical quality of new sources
  - Issue statistical “best practices”
  - Use non-traditional sources to complement, augment and/or replace existing series
- But most NSOs have limited to no experience in working with big data
- Innovation will often come from third party organizations

# Who are the stakeholders for the new data literacy that is required?

- National Statistical Organizations (NSOs)
- Regulators (especially telecom regulators)
- Private sector
- Civil society
- Academics and researchers (especially for extending the state of the art)



# What's the role for the UN?

- There is already a UN initiative working on big data: UN Global Pulse
  - Works as a network of innovation labs (New York, Jakarta, and Kampala so far)
  - Partners with experts from UN agencies, governments, academia, and the private sector
- What will the role be?
  - Inform and catalyze the discussion on global best practices and the use of big data for development
  - Look for non-traditional innovators within countries and facilitate new partnerships and collaborations

# Thank You.

Sriganesh Lokanathan

Team Leader – Big Data Research, LIRNEasia

[sriganesh@lirneasia.net](mailto:sriganesh@lirneasia.net)

<http://lirneasia.net/projects/bd4d/>