Big data for development research in the Global South: Experiential lessons

LIRNE*asia*

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

Oxford Internet Institute Symposium on Big Data and Human Development, September 15-16, 2016

LIRNE*asia*

Pro-poor. Pro-market.

Some development problems of interest to LIRNEasia . . .

- More people live in cities than in rural areas since 2008
  - How can we make cities more livable?
  - Is there a role of ICTs, not just more roads, transit, etc.?
- Infectious diseases are posing threats
  - Can we make better decisions re allocating scarce resources?
- Governments are flying blind, with ineffective National Statistical Organizations unable to give timely data needed to better target expenditures, assess programs or achieve SDGs
  - Are there ways to remedy this?

Comprehensive coverage of population needed. Sources of data?

- 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 at this time

Country	Mobile SIMs/100	Internet users/100	Facebook users/100
	2015	2014	2016
Myanmar	69	2	20.4
Bangladesh	82	10	13.7
Pakistan	67	14	14.3
India	76	18	11.5
Sri Lanka	125	26	19.6
Philippines	120	40	54.6
Indonesia	131	17	33.8
Thailand	125	35	58.9

Sources: https://www.gsmaintelligence.com/;
http://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2015/MISR2015-w5.pdf;
Facebook advertising portal; http://data.worldbank.org/indicator/SP.POP.TOTL

LIRNE*asia*
Pro-poor. Pro-market

4

Data used in the research

- Multiple mobile operators in Sri Lanka provided four different types of meta-data
  - Call Detail Records (CDRs)
    - Records of calls
    - SMS
    - Internet access
  - Airtime recharge records
  - No Visitor Location Registry (VLR) data, because they are written over & not stored
- Data sets do not include any Personally Identifiable Information
  - All phone numbers are pseudonymized
  - LIRNE*asia* does not maintain any mappings of identifiers to original phone numbers
- Historical, not real time; therefore analyzed in batch mode in a hardware stack costing < USD 30k
- Cover 50-60% of users; very high coverage in Western (where Colombo the capital city in located) & Northern (most affected by civil conflict) Provinces, based on correlation with census data

LIRNEasia
Pro-poor. Pro-market

EXAMPLE: POPULATION DENSITY & MOVEMENT

Population density changes in Colombo region: weekday/ weekend

Pictures depict the change in population density at a particular time relative to midnight
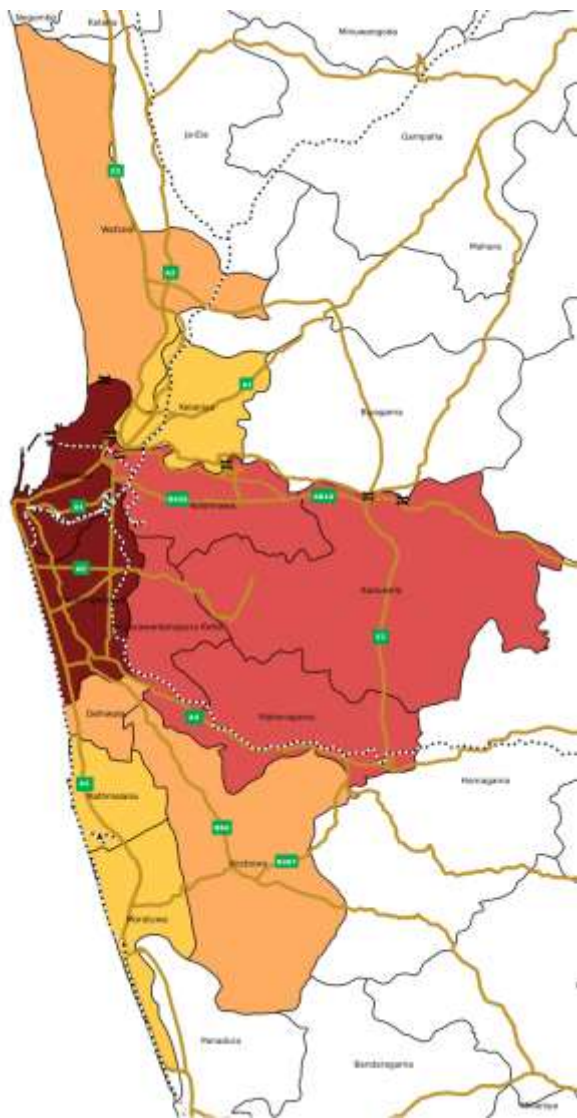
Our findings closely match results from expensive & infrequent transportation surveys; are cheaper & can be produced as needed
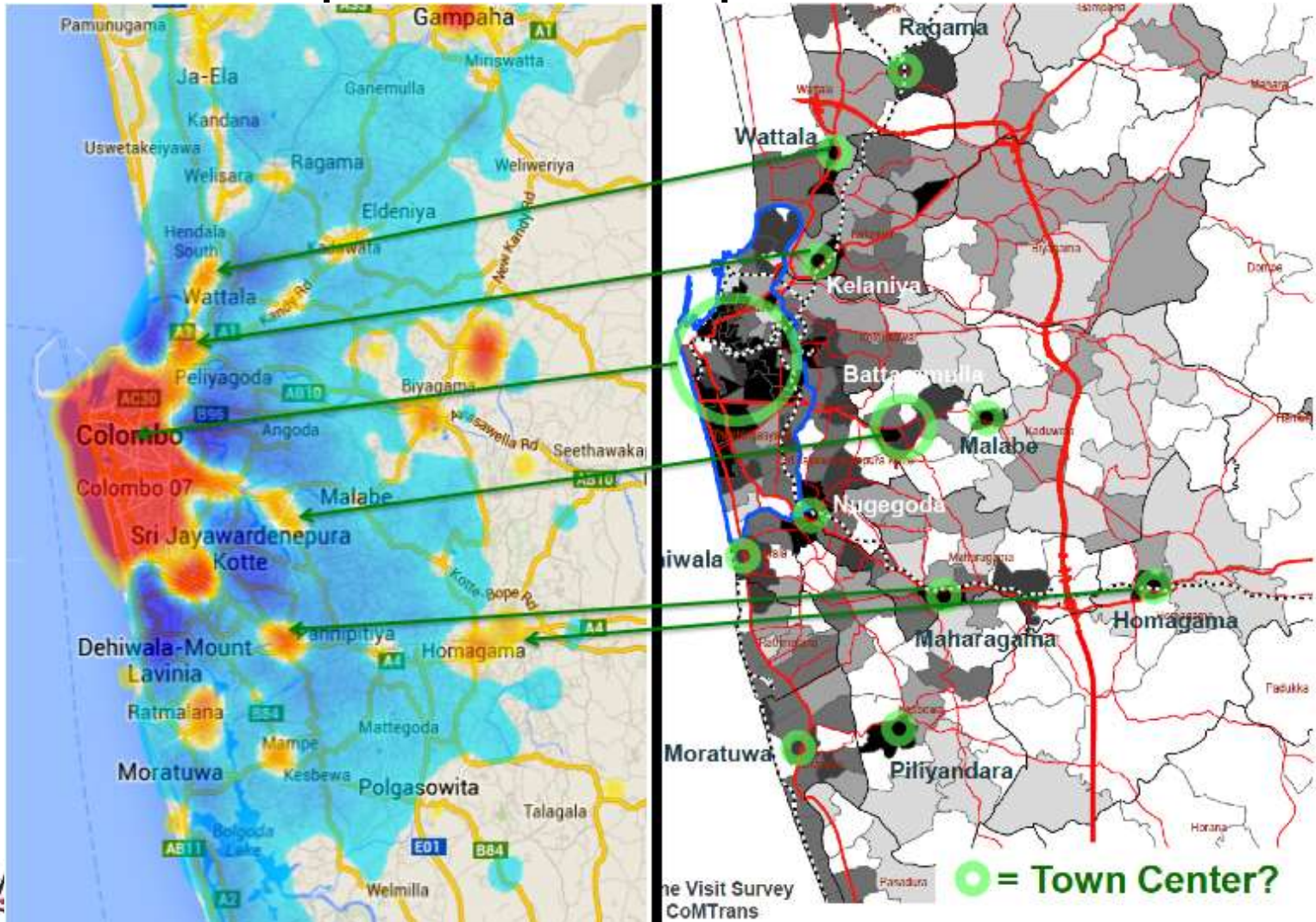
46.9% of city's daytime population comes from outside. Potential configurations of a Metropolitan Corporation

Home DSD	Population	Percentage contribution to Colombo's daytime population
Colombo City (2 DSDs)	555,031	53.1
Maharagama	195,355	3.7
Kolonnawa	190,817	3.5
Kaduwela	252,057	3.3
Sri J'pura Kotte	107,508	2.9
Dehiwala	1,387,834	62.6
Kesbewa	244,062	2.5
Wattala	174,336	2.5
Kelaniya	1,807,600	74.1
Ratmalana	95,162	2.0
Moratuwa	167,160	1.8
Total	2,204,015	79.9

CHALLENGE 1: POLICY IMPACT

From supply-push to demand-pull

	Event
2014 Oct	• Founding Chair has one-on-one meeting with Secretary, Urban Development ➔ presentation to urban development professionals in November 2014 agreed upon, but postponed due to early announcement of Presidential Election
2015 Jan	• Big Data team conducts a public lecture organized by Institute of Engineers Sri Lanka (IESL)
2015 Feb	• Email contact made with new DG of UDA with offer to brief on LA's ongoing research (meetings planned but don't happen) • First media interactions in Sri Lanka
2015 May	• Big Data Team Leader makes 5-minute presentation at Workshop on implementation of transportation master plan for Ministry of Internal Transport organized by UoM's Dept. of Transport & Logistics. Attended by domain specialists, academics, researchers, officials from UDA, RDA, etc.
2015 Jun - Sept	• Big Data Team Leader invited to join planning team for the Western Region Megapolis Planning Project to provide insights from MNBD. Our name suggested by one of the attendees in the May 2015 event, who was appointed to lead one of the committees working on the WRMPP • In total 9 meetings were attended, culminating in a presentation to all the committees on insights from LIRNEasia big data research related to urban and transportation planning
2015 Aug	• Big Data Team Leader presents at Workshop on Integrated Land Use Transport Modeling Practices in Sri Lanka and around the World organized by UoM's Transportation Engineering Division. Attended by domain specialists, academics, researchers, officials from UDA, RDA, etc.

Policy Enlightenment / Supply-push

Demand-Pull

Sri Lanka's highest-circulation English newspaper

Sunday Times 2   Sunday April, 12, 2015

Big data can make South Asian cities smarter

Big data analysis can help citizens make smart choices and plan cities, says Nalaka Gunawardene
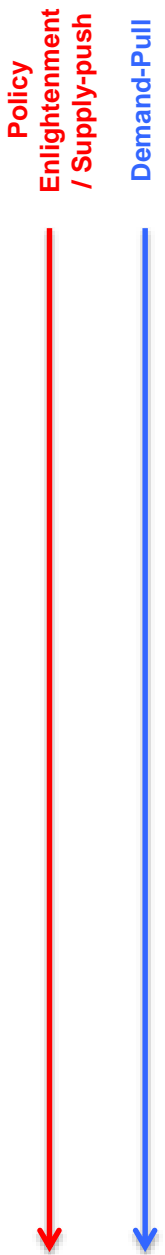
View(s):

With more than half of humanity living in cities, there is an urgent need to improve urban planning, design and management.
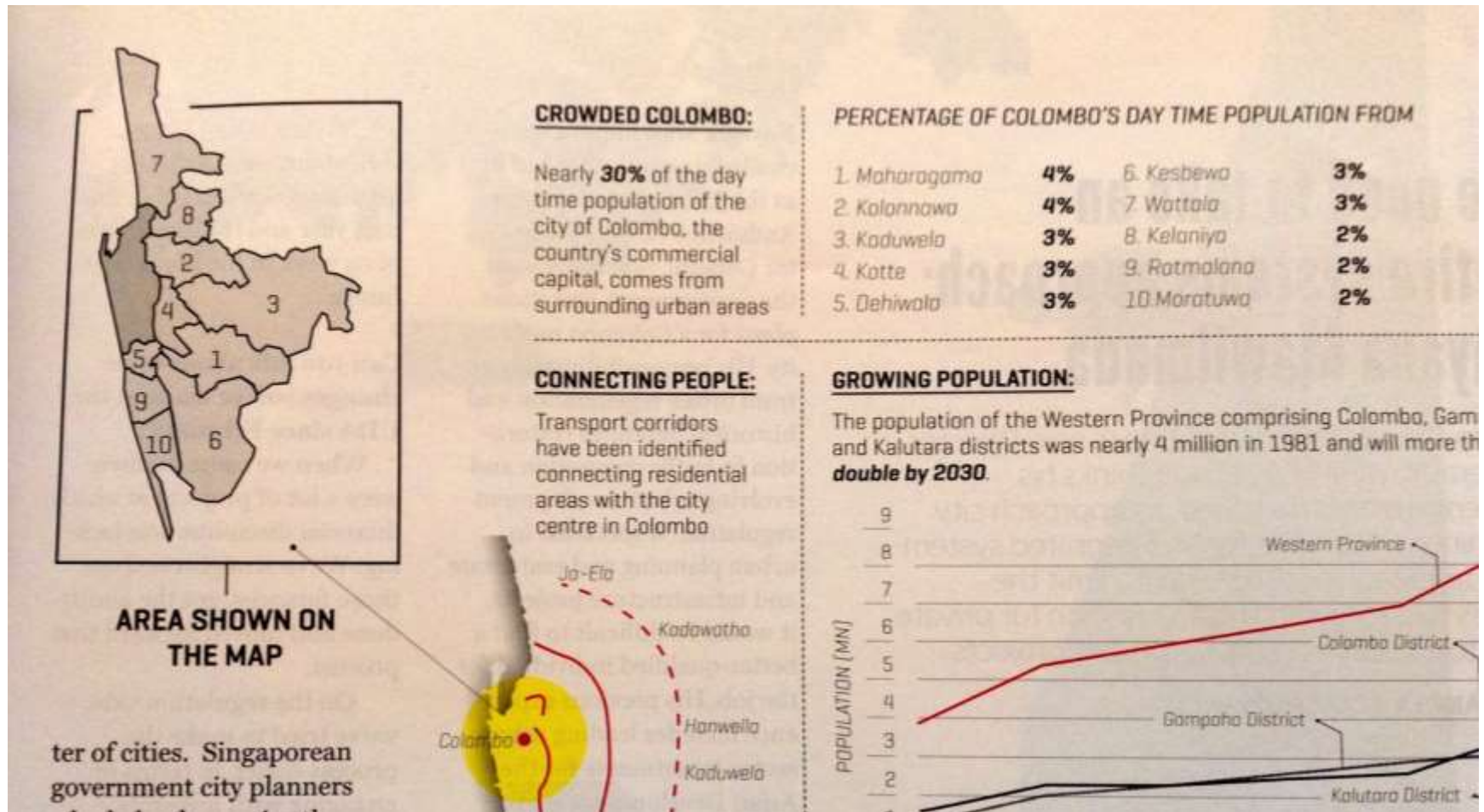
Until now, policy makers and planners have struggled to keep up with trends. Changes were too fast, and variables too dynamic, for conventional surveys or censuses to capture them adequately.

From supply-push to demand-pull

	Event
2015 Sept	• Big Data Team Leader attends launch of World Bank Report on "Leveraging Urbanization in South Asia" • Side discussions with DG of UDA on LIRNEasia's ongoing research leads to one-on-one meeting with DG the following day to brief him on LIRNEasia research
2015 Dec	• UDA, UDA's Professionals Association, & Young Planners Forum of Institute of Town Planners, Sri Lanka organize special session for LA to present ongoing research. • Echelon Magazine report on Megapolis plans include charts given by UDA on source locations of Colombo's daytime population developed by LIRNEasia (without acknowledgement)
2016 Jan	• Big Data team invited to make presentation to Sri Lanka Strategic Cities Development Project working on Kandy
2016 Feb	• DG UDA requests additional mobility and land-use insights on Kandy
2016 Feb	• Sri Lanka Strategic Cities Development Project reaches out to LIRNEasia for insights on foot traffic in Kandy. Our data not suitable but we brainstorm possible methodologies
2016 Aug	• UDA requests additional finer-grained mobility insights for specific areas in Western Province

Policy Enlightenment / Supply-push

Demand-Pull

Pro-poor. Pro-market

13

Western Regional Megapolis Project (WRMP)

CROWDED COLOMBO:

Nearly **30%** of the day time population of the city of Colombo, the country's commercial capital, comes from surrounding urban areas

PERCENTAGE OF COLOMBO'S DAY TIME POPULATION FROM

1. Maharagama	4%	6. Kesbewa	3%
2. Kolonnawa	4%	7. Wattala	3%
3. Kaduwela	3%	8. Kelaniya	2%
4. Kotte	3%	9. Ratmalana	2%
5. Dehiwala	3%	10. Moratuwa	2%

CONNECTING PEOPLE:

Transport corridors have been identified connecting residential areas with the city centre in Colombo

GROWING POPULATION:

The population of the Western Province comprising Colombo, Gam and Kalutara districts was nearly 4 million in 1981 and will more th **double by 2030**.

AREA SHOWN ON THE MAP

ter of cities. Singaporean government city planners

Source: Interview with Western Region Megapolis Authority in Echelon magazine (December 2015, pp. 63)

LIRNEasia
Pro-poor. Pro-market

14

Lessons

- No demand for insights when we started in 2012; but had we not started on the research when we did, no results would have been available when the policy window opened

- Given difficulty of assessing quality of big-data research, the credibility of LIRNEasia has been of value
  - More work needs to be done to make potential users of this research more knowledgeable

- Supply-push approach helped create the conditions for essential demand-pull
  - However, political changes and appointments which were outside our control were critical in creating demand

LIRNEasia
Pro-poor, Pro-market

CHALLENGE 2: HUMAN RESOURCES

Data scientists are in short supply; Needed are multi-disciplinary teams

- We prioritize analytical thinking over knowledge of big data tools in our recruitment interviews

- Team members have different specialties
  - Staff and collaborators have a mix of computer-science skills, statistics, and domain knowledge

Staff

- Danaja Maldeniya (Computer Science, Statistics)
  - Now at U of Michigan but still collaborating
- CD Athuraliya (Computer Science)
- Dedunu Dhananjaya (Software Engineer, Systems Administrator)
  - Moved to private sector, but still collaborating
- Isuru Jayasooriya (Computer Science, Machine Vision)
- Kaushalya Madhawa (Computer Science, Statistics)
  - Now at Tokyo Institute of Technology
- Keshan de Silva (Computer Science, Agent Based Simulations)
- Lasantha Fernando (Computer Science)
- Madhushi Bandara (Computer Science)
  - Now at U of New South Wales but still collaborating
- Nisansa de Silva (Computer Science)
  - Now at U of Oregon, but still collaborating
- Robert Galyean (Mathematics, Physics)
- Prof. Rohan Samarajiva (Public Policy)
- Sriganesh Lokanathan (Public Policy, Computer Science)
- Shazna Zuhyle (Research Manager)
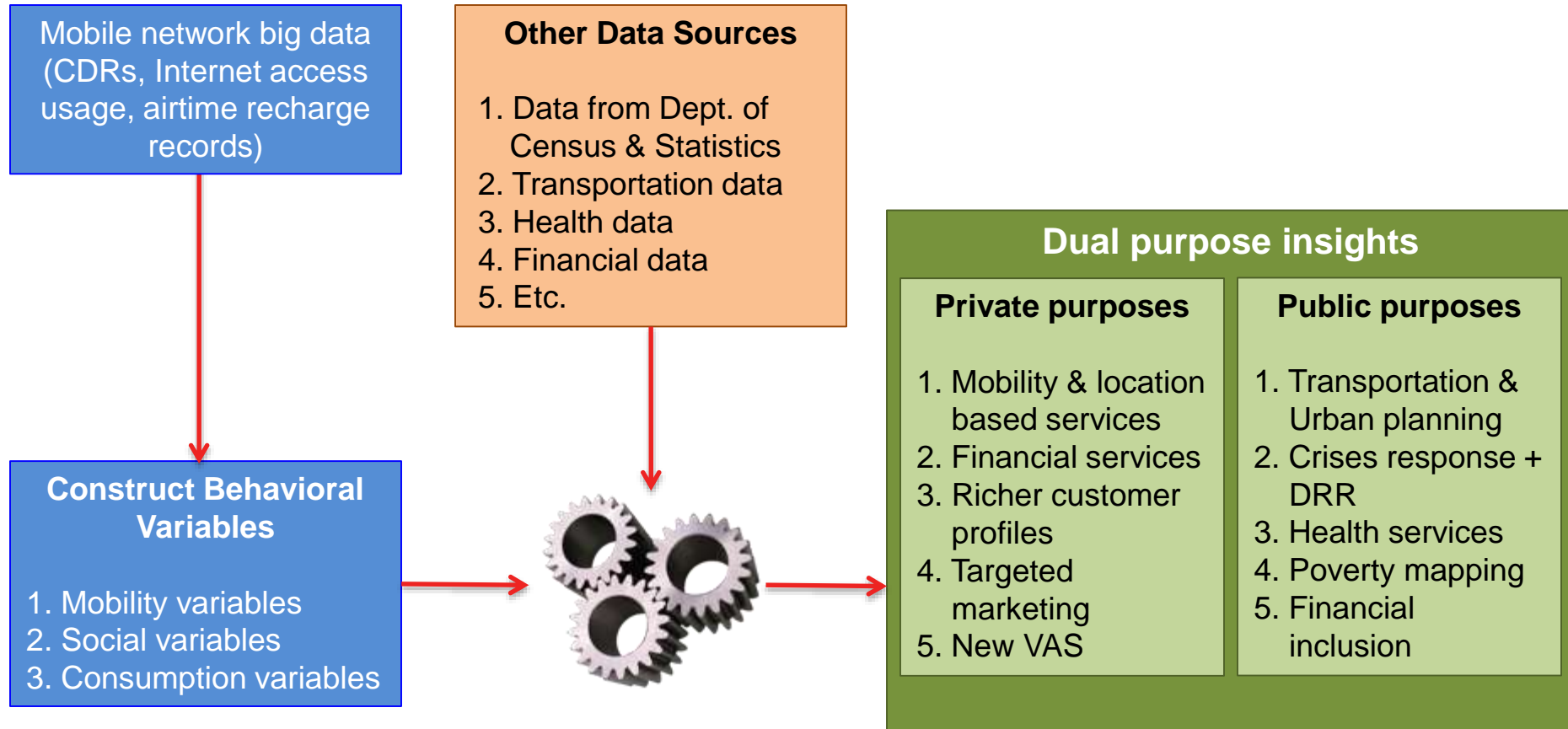- Thavisha Gomez (Research Manager)

Collaborators

- Prof. Amal Kumarage (Dept. of Transport & Logistics, UoM)
  - Transportation, Urban Planning
- Dr Amal Shehan Perera (Dept. of CSE, UoM)
  - Data Mining
- Fields of View
  - Indian research institute specializing in games and simulations for public policy issues
- Gabriel Kreindler (MIT)
  - Economics, Statistics
- Prof Joshua Blumenstock (UC Berkley, School of Information)
  - Data Science, Economics, Statistics
- Prof Moinul Zaber (U of Dhaka)
  - Data Science, Public Policy
- Shibasaki & Sekimoto Laboratory, University of Tokyo
  - Big Data for Development research practice
- Yuhei Miyauchi (MIT)
  - Economics, Statistics

Flow through: We're always hiring

- A computer-science graduate does not think of working in a think-tank; often looking to join a software firm
  - LIRNE*asia* has done presentations in public forums and at universities to broaden horizons
- Key selling point is the work that we do and our partnerships
  - Rewarding to see research being used
  - Good opportunities for publication and conference papers
  - LIRNE*asia* encourages individual researchers to build their "brands"
  - Ideal for admission into good PhD programs; current funded placements
    - U Oregon
    - U Michigan
    - Tokyo Institute of Technology
    - U New South Wales

LIRNE*asia*
Pro-poor. Pro-market

CHALLENGE 3: DATA

Mobile network big data + other data ➔ rich, timely insights that serve private **as well** as public purposes

Mobile network big data (CDRs, Internet access usage, airtime recharge records)

Other Data Sources

1. Data from Dept. of Census & Statistics
2. Transportation data
3. Health data
4. Financial data
5. Etc.

Construct Behavioral Variables

1. Mobility variables
2. Social variables
3. Consumption variables

Dual purpose insights

Private purposes

1. Mobility & location based services
2. Financial services
3. Richer customer profiles
4. Targeted marketing
5. New VAS

Public purposes

1. Transportation & Urban planning
2. Crises response + DRR
3. Health services
4. Poverty mapping
5. Financial inclusion

Mobile network big data + other data → rich, timely insights
serve private **as well** as public purposes

Private sector data

Government data

Mobile network big data (CDRs, Internet access usage, airtime recharge records)

Other Data Sources

1. Data from Dept. of Census & Statistics
2. Transportation data
3. Health data
4. Financial data
5. Etc.

Dual purpose insights

Private purposes	**Public purposes**
1. Mobility & location based services	1. Transportation & Urban planning
2. Financial services	2. Crises response + DRR
3. Richer customer profiles	3. Health services
4. Targeted marketing	4. Poverty mapping
5. New VAS	5. Financial inclusion

Construct Behavioral Variables

1. Mobility variables
2. Social variables
3. Consumption variables

Mobile network big data

- No established process exists, therefore prior relationships matter
- Basic process
  1. Obtain in-principle agreement from CEOs of companies, ideally at least two
  2. Painfully negotiate specifics with 2nd and 3rd tier management (~ 6 months)
  3. Approach other operators
- Throughout, mutual benefits emphasized
  - Methods for deriving public policy insights can also be adapted for commercial purposes

LIRNEasia
Pro-poor. Pro-market

Responding to operator concerns

- Will the regulator object?
  - The specific data requested did not contravene existing laws or license conditions
  - Data we obtain is pseudonymized with no links to original numbers

- Will this research reveal any proprietary business intelligence?
  - We combine data from multiple operators
  - All researchers sign Non Disclosure Agreements with LIRNEasia
  - Operators sign off on our results before public release

LIRNEasia
Pro-poor. Pro-market

How can access to data for research for public purposes be made easier?

- Reduce transaction costs:
  - Standardized agreement template(s)
- Pro-actively deal with privacy and competition concerns
  - LIRNEasia working with operators in the region on self-regulatory guidelines for minimizing potential harms
    - Draft self-regulatory guidelines developed by LIRNEasia available at http://lirneasia.net/2014/08/what-does-big-data-say-about-sri-lanka/
- New operational models with third-party data guardians?
  - E.g., Yale University Open Data Access (YODA) project acting as data guardians and providing researchers access to Johnson & Johnson clinical trial data

LIRNEasia
Pro-poor. Pro-market

26

Government data are not any easier

- Some processes exist for getting data, but no uniform procedures
  - Easier for government-affiliated entities; LIRNEasia is not one
- Different, not mutually exclusive, approaches
  - Submitting a data request proposal
  - Meeting with senior officials
  - Spending time talking to junior officials and negotiating access
  - Building compelling story based on mobile network big data as part of rationale for access to government data
  - Partnering with government-affiliated organization (e.g., University of Moratuwa)