Confluence (or lack thereof) of data analytics and law

Rohan Samarajiva, Ph.D., Attorney at Law

SAARClaw 2017, Colombo, 28 October 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 and the Department for International Development UK . Also supported by New Venture Fund.

UKaid

from the British people

DATA ANALYTICS FOR PUBLIC PURPOSES (I) URBAN PLANNING

Some development problems that may be addressed using big data . . .

- Worldwide, 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 development goals such as SDGs
  - Are there ways to remedy this?

LIRNEasia
Pro-poor Pro-market

Comprehensive coverage of population needed for most public-policy problems. 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
- In some cities, electricity billing data also has comprehensive coverage

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

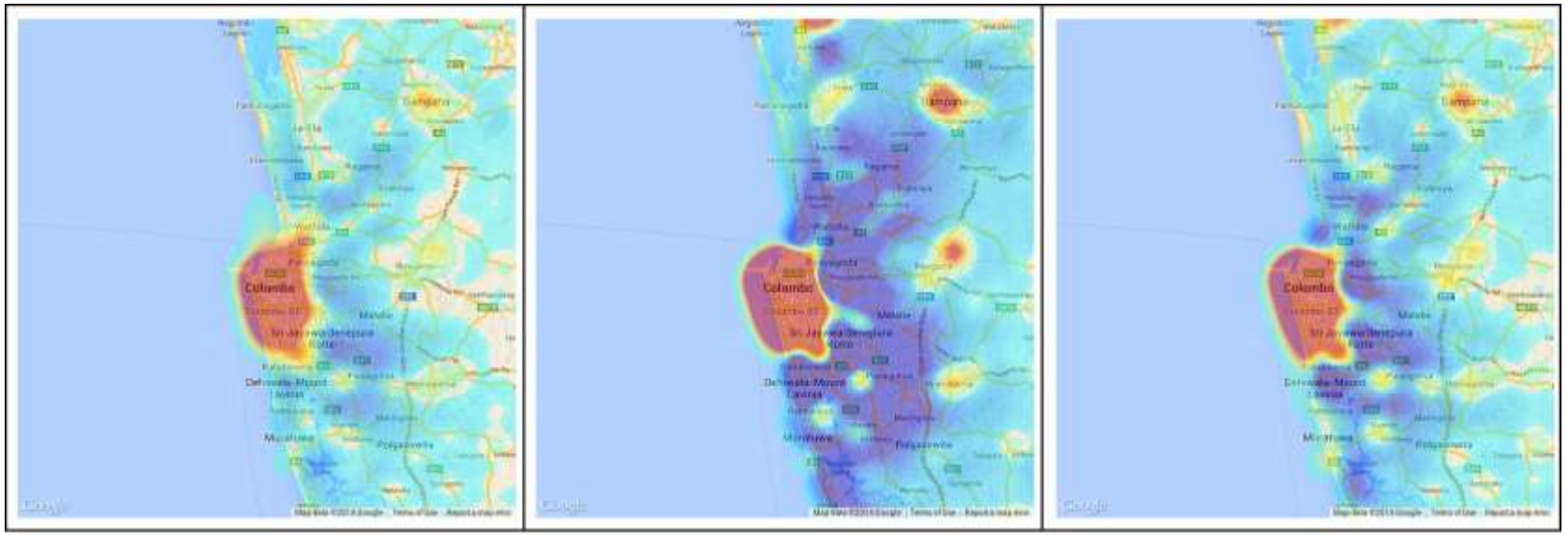
Data used in the research

- 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

Population density changes in Colombo region: weekday/ weekend

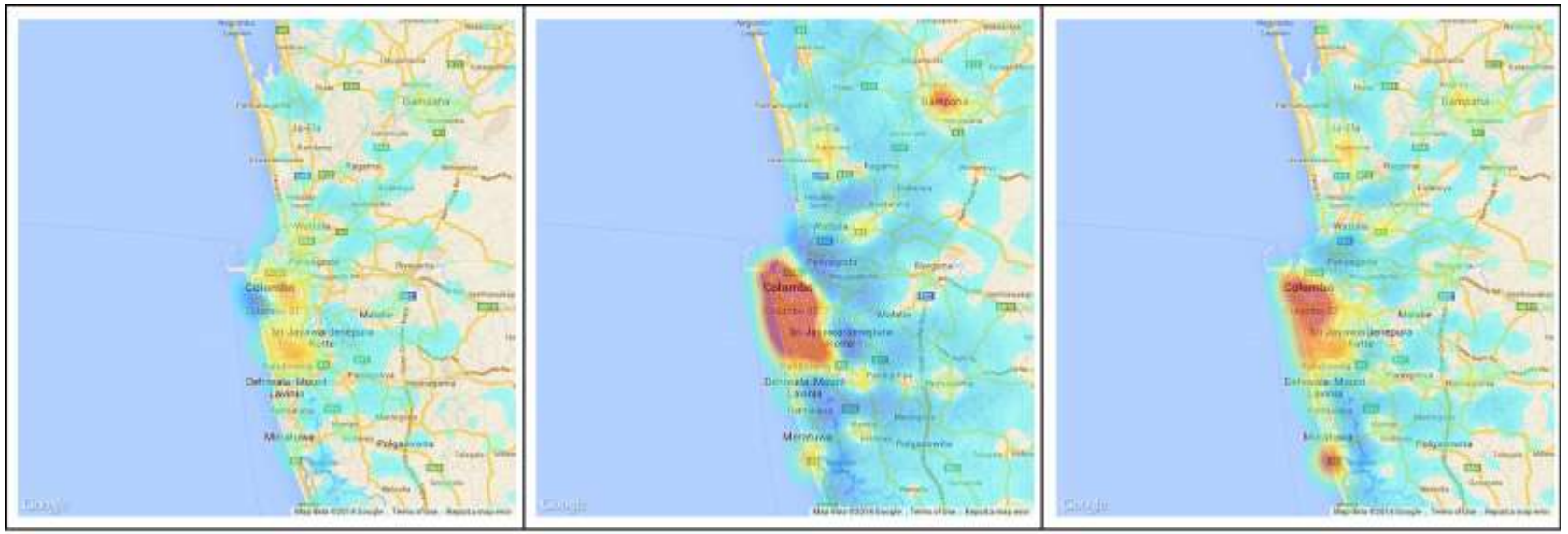
Pictures depict the change in population density at a particular time relative to midnight

Weekday



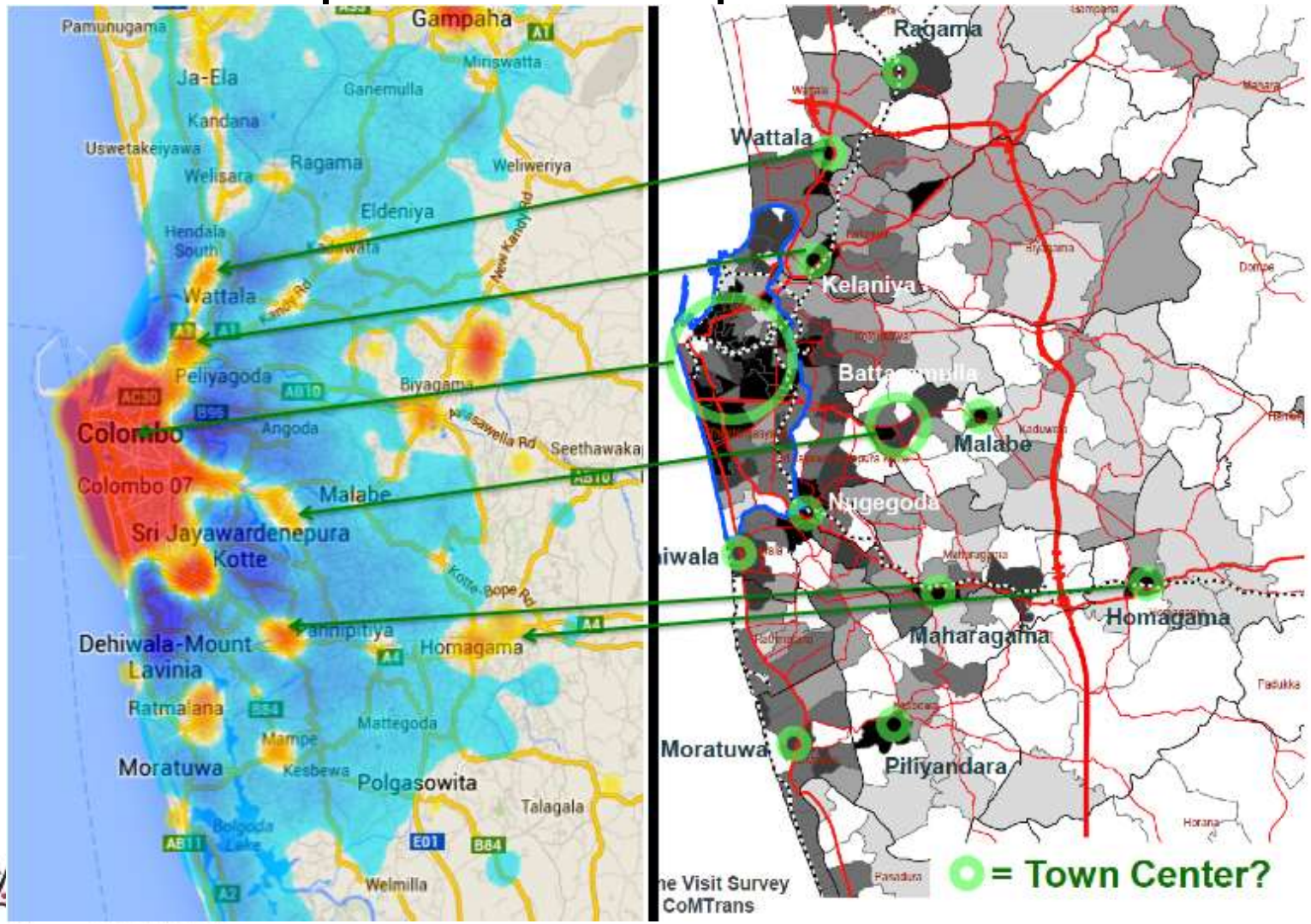
Time 06:30 Time 12:30 Time 18:30

Sunday

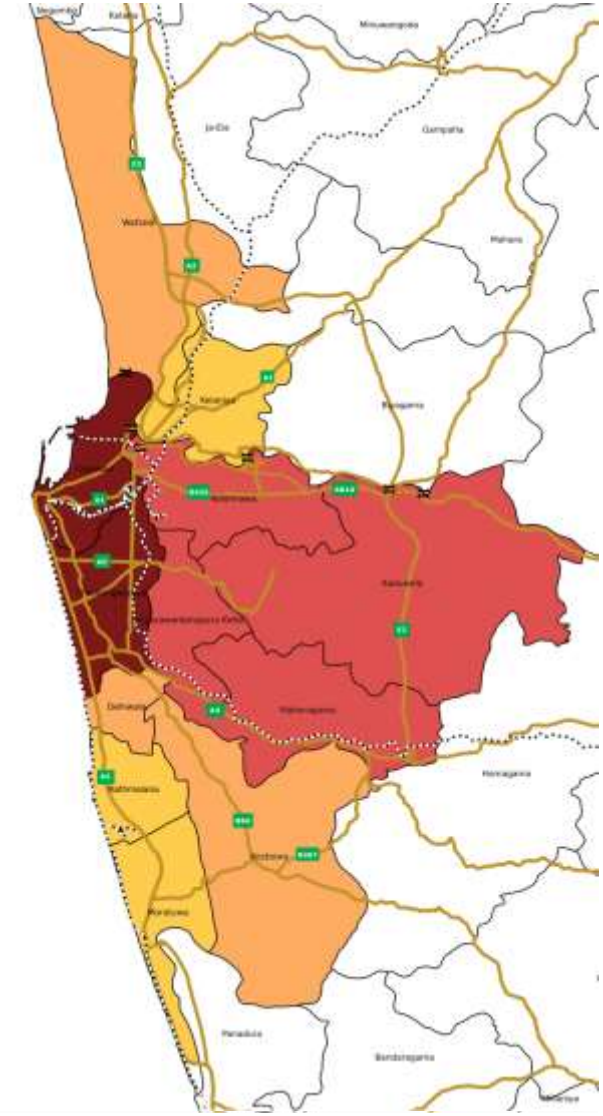


Decrease in Density Increase in Density

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	195,855	53.7
Kolonnawa	190,817	3.5
Kaduwela	252,057	3.3
Sri J'pura Kotte	107,508	2.9
Dehiwala	87,884	2.6
Kesbewa	244,062	2.5
Wattala	174,336	2.5
Kolaniya	197,690	2.1
Ratmalana	95,162	2.0
Moratuwa	167,160	1.8
Total	2,204,015	79.9

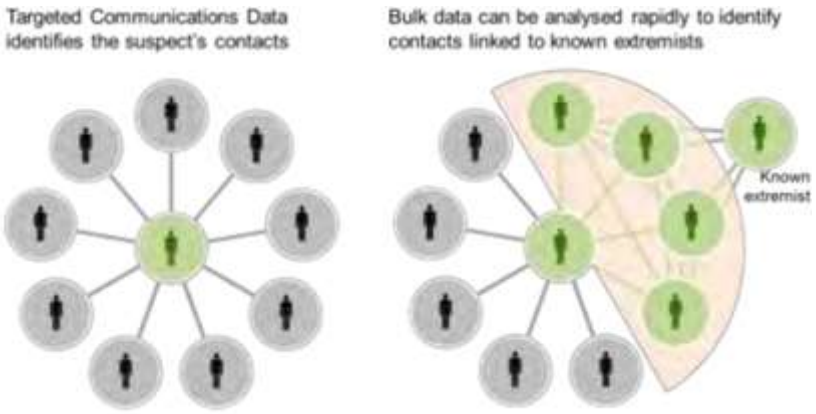
DATA ANALYTICS FOR PUBLIC PURPOSES (2): DETERRING/INVESTIGATING TERRORIST ACTS

Claim of deterring terrorist acts through analysis of mobile network big data

- “In 2010, a network of terrorists – comprising groups in Cardiff, London and Stoke-on-Trent - planned a series of bomb attacks at several symbolic locations in the UK, including the London Stock Exchange. Complex analysis of bulk acquisition data played a key role in identifying the network. The task was made particularly challenging by the geographical separation of the groups. Nine members of the network were subsequently charged and pleaded guilty to terrorism offences relating to the plot. Eight members of the network pleaded guilty to engaging in conduct in preparation for acts of terrorism.”
- According to MI5, bulk communications data acquisition is: "who", "where", "when", "how" and "with whom" of communications, but not what was written or said. https://www.mi5.gov.uk/bulk-data”

Anderson, D. (2016). Report of the Bulk Powers Review. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/546925/56730_Cm9326_WEB.PDF

Methods used, in outline



- No pseudonymization
- Start with a one suspect, based on conventional investigative techniques
- Then use first circle of communication
- Move on to second circle

Operational Case for Bulk Powers. (n.d.). Retrieved from www.gov.uk/government/uploads/system/uploads/attachment_data/file/504187/Operational_Case_for_Bulk_Powers.pdf

Use of MNBD in investigations

- "The identification of persons of interest itself may entail casting a wider net. For instance, if a crime was committed at a particular place, at a particular time, identifying possible suspects may mean first identifying all the people who were in the vicinity at that particular time (for example, using a cell tower dump). The analysis of call detail records can provide important insights into the mobility patterns and social networks of individuals, and provide vital information that could show important connections that solve crimes by helping to show what suspects were before, during and after a crime was committed."

India's Central Monitoring System (CMS)

- All service providers in India required to have Lawful Interception Systems at their premises in order to carry out targeted surveillance of individuals by monitoring communications running through their networks
- Now, all TSPs in India are required to integrate Interception Store & Forward (ISF) servers with their pre-existing Lawful Interception Systems. ISF servers installed in the premises of TSPs are connected to the Regional Monitoring Centres (RMC) of the CMS. Each RMC is connected to the CMS)
- Not only can the CMS authority have centralized access to all data intercepted by TSPs all over India, but that the authority can also bypass service providers in gaining such access

http://www.dot.gov.in/sites/default/files/Annual%20Report%202016-17.pdf?download=1

6/C

Government of India
Ministry of Communications and IT
Department of Telecommunications
(Access Service Cell)
Sanchar Bhawan, 20, Ashok Road, New Delhi-110001

File No: 800-12/2013-AS.II **Dated:** ---June' 2013

To
All UAS Licensee(s)

AMENDMENT 2 OF 2013

Subject: Amendment to the UAS License agreement regarding Central Monitoring System.

The Government has decided to set up Centralized Monitoring System (CMS) for lawful interception and monitoring of communications. For the implementation of the same, LICENSEE's Lawful Interception System needs to be connected to the CMS at Regional Monitoring Centre (RMC) through Interception Store and Forward (ISF) server placed in LICENSEE's premises.

For this purpose, kindly find hereby enclosed the amendment to the condition 41.10 of the UAS license(s)

(A.K.Tirkey)
AD (AS-II)

Copy to:
1. Secretary, TRAI
2. Sr. DDG, TEC
3. Sr. DDG (TERM), DoT
4. DDG (Security), DoT
5. All DDsG TERM.
6. Director (AS-I)/ Director (AS-III)/ Director (AS-IV), Dir(AS-V), DoT

Nepal's legal developments

- Babu Ram Aryal et al vs Government of Nepal et al. (2016)
  - Justice Ran Bahadur Bam was shot dead on 31st May 2012
  - CDR of 500k calls & 60k SMS collected but no useful information obtained
  - CDRs retained
  - News reports were published 27 August 2012 about possible misuse
  - Public Interest Litigation initiated at Supreme Court of Nepal 6 September, 2012; judgment 4 February 2016

Nepal Supreme Court decision

- Retention of CDR and SMS by security agency is illegal
- Institutions or Departments which store (retain) the information or are custodians of the information cannot make use of them at their discretion
- Inability to protect information in their custody by fear or any influence is unlawful
- Implementation of Article 28 (Right to Privacy) is mandatory; concerned government agency is directed to make a necessary law adopting the universal principle of the privacy
- Until legal arrangements are made, if an urgent situation arises in the course of criminal investigation to collect such information, pursuant to any law, it is directed to receive mandatory approval from a district court

China’s ambitious plan

- Imagine a world where many of your daily activities were constantly monitored and evaluated: what you buy at the shops and online; where you are at any given time; who your friends are and how you interact with them; how many hours you spend watching content or playing video games; and what bills and taxes you pay (or not). It's not hard to picture, because most of that already happens, thanks to all those data-collecting behemoths like Google, Facebook and Instagram or health-tracking apps such as Fitbit. But now imagine a system where all these behaviours are rated as either positive or negative and distilled into a single number, according to rules set by the government. That would create your Citizen Score and it would tell everyone whether or not you were trustworthy. Plus, your rating would be publicly ranked against that of the entire population and used to determine your eligibility for a mortgage or a job, where your children can go to school - or even just your chances of getting a date.

http://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion

Privacy-associated harms

- As commonly understood, “is a sweeping concept, encompassing (among other things) freedom of thought, control over one’s body, solitude in one’s home, control over personal information, freedom from surveillance, protection of one’s reputation, and protection from searches and interrogations” (Solove, 2008, p. 1)
- Attempts to define privacy in terms of boundary control by individuals (e.g., Samarajiva, 1994: 90) are difficult to translate into practical policy

Our approach

- Focus on harms as identified by Solove's research that fall into four general types
  - Information collection;
  - Information processing;
  - Dissemination of information; and
  - Invasion
- Develop remedies to prevent harms
- Work with all stakeholders to operationalize safeguards

16 harms within umbrella meaning ➔ 9 of relevance to MNBD

1. Surveillance, interrogation (1/2)
2. Aggregation, identification, insecurity, secondary use, exclusion (5/5)
3. Breach of confidentiality, disclosure, exposure, increased accessibility, blackmail, appropriation, distortion (3/7)
4. Intrusion, decisional interference (0/2)

Most from information processing cluster; next from dissemination

LIRNEasia
Pro-poor Pro-market

Examples of reasons for inclusion/exclusion

- Surveillance v ~~interrogation~~
  - Surveillance is obviously relevant
  - Interrogation is the pressuring of individuals to divulge information (physical coercion, not about information)
- Disclosure v ~~exposure~~
  - Both involve dissemination of true information, but exposure is limited to information about body and health
- ~~Decisional interference~~
  - "Right to privacy" is some countries encompasses a woman's decision whether or not to terminate pregnancy

Remedy	Identified potential harm	Include in agreements transferring identifiable MNBD	Include in agreements transferring anonymized MNBD
Mobile Network Operators (MNOs) will not engage in active surveillance of their customers, except as required by applicable law. MNOs will desist from collecting more data than are needed for the efficient operation of the networks and the supply of good service to customers. To the extent feasible, data collection practices will be transparent.	***Active surveillance***	No. Applying only to MNOs, this need not be included in agreements. However, active surveillance is a root cause of problems that could be manifested in other forms at the subsequent information processing and dissemination phases.	No. Applying only to MNOs, this need not be included in agreements. However, active surveillance is a root cause of problems that could be manifested in other forms at the subsequent information processing and dissemination phases.

Remedy	Identified potential harm	Include in agreements transferring identifiable MNBD	Include in agreements transferring anonymized MNBD
Any agreement transferring identifiable data to a third party will also transfer responsibility to maintain safeguards to ensure security of individually identifiable data.	***Insecurity***	Yes	No

Remedy	Identified potential harm	Include in agreements transferring identifiable MNBD	Include in agreements transferring anonymized MNBD
The agreement governing the transfer will include provisions to minimize risks posed by increased accessibility when data are released to third parties.	***Increased accessibility***	Yes	Yes

Group harms

Example: Socio-economic mapping

- Governments/IGOs wish to identify the poor so services may be efficiently delivered to them.
- Today, socio-economic mapping seeks to literally map or associate poverty on spatial representations. In future, may be extended beyond mapping in the literal sense. The analogy is to the zip-code-based voter mobilization efforts of past US elections versus the current precision-targeted get-out-the-vote exercises.

If the poor can be identified, so can the rich

- Will this result in prioritization of areas where the rich live in terms of service delivery?
- In competitive markets, suppliers are not expected to serve the entire market at the very outset or even at any point. Uncertainty about demand is normal. Therefore, suppliers enter in limited geographical areas or focus on particular market segments at the outset. It is only on the basis of feedback from these activities that the firm will scale up. Some firms will adopt niche strategies and never seek to serve the entire market.

Dangers of safeguarding against group harms, by creating new right of group privacy

- Rights are usually understood to belong to individuals, not to groups. The only group right recognized in international law is that of peoples having the right of self-determination
- Prejudice against actions based on group attributes would pretty much put an end to efforts to improve the functioning of society in systematic, evidence-based ways, e.g.,
  - Routine to associate various characteristics or behaviors with persons living in geographical areas (e.g., in poverty mapping), by age group and gender and so on
  - Desirable to "target" various policy measures to specific groups and indeed to improve the targeting by various means.
- Without group identification it will be impossible for modern societies to function