11th World Telecommunication/ICT Indicators Symposium (WTIS-13)

Mexico City, México, 4-6 December 2013

Contribution to WTIS-13

Document C/27-E
9 December 2013

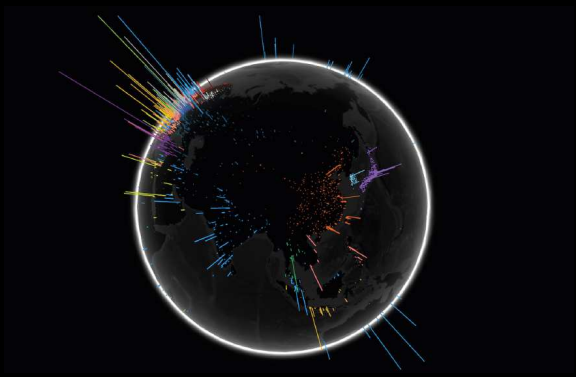
English

SOURCE: UN Global Pulse

TITLE: Big Data in Real Time: Toward a New Evidence Base for Impact

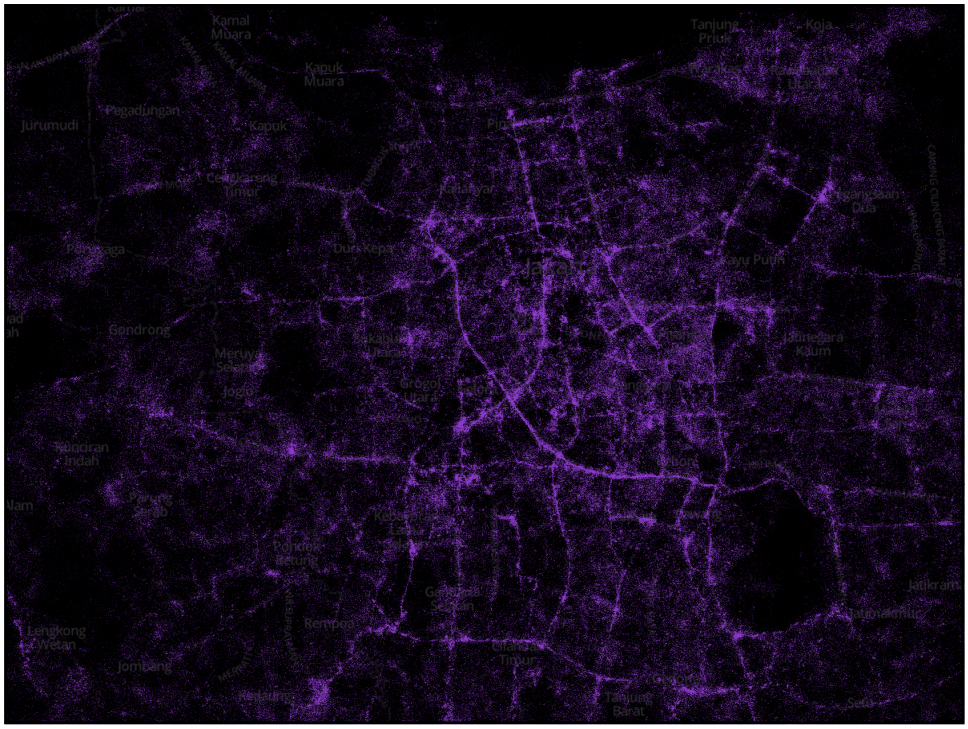
Big Data in Real Time
Toward a New Evidence Base for Impact

Robert Kirkpatrick
Director, UN Global Pulse

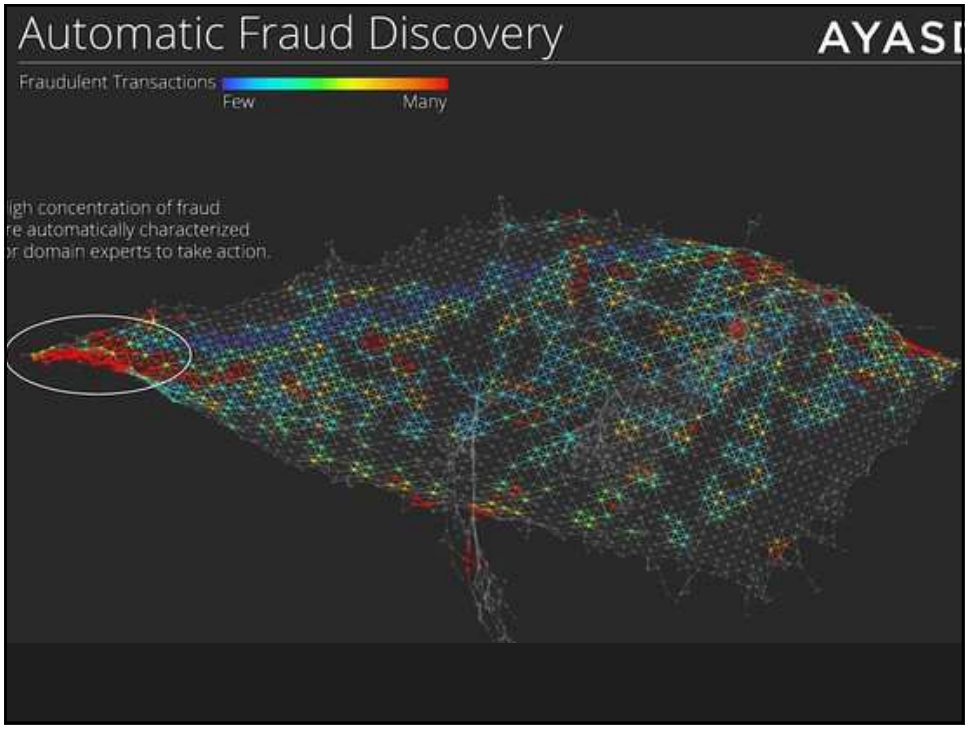


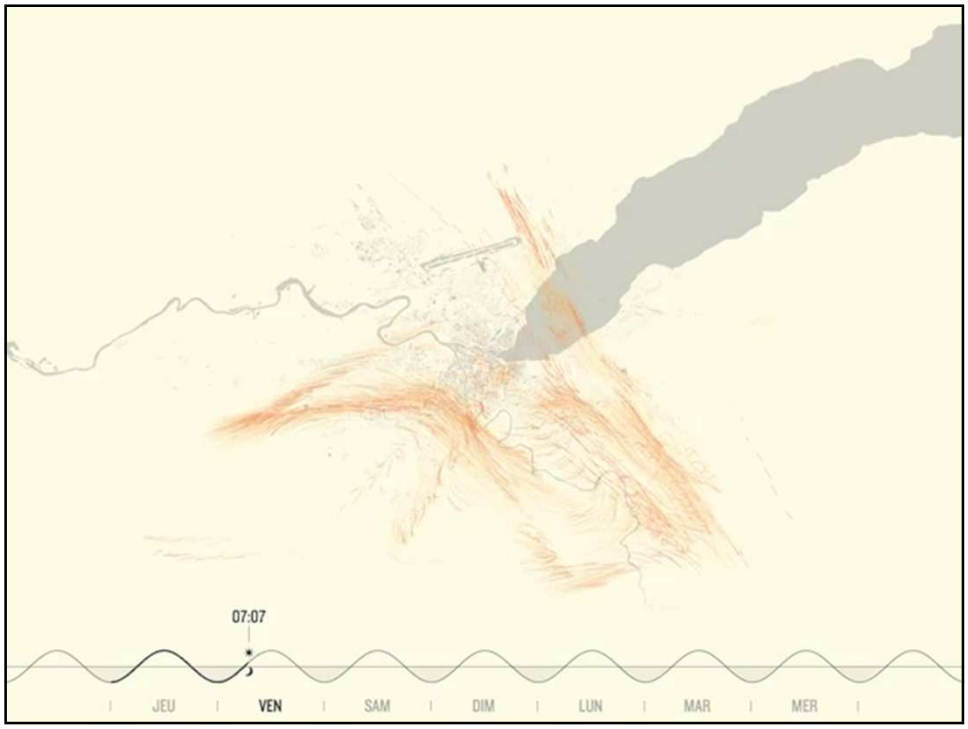
GLOBAL PULSE

www.unglobalpulse.org @unglobalpulse


Kamal Muara
Kapuk Muara
Tanjung Priuk
Koja
Pegadungan
Jurumudi
Kapuk
Cengkareng Timur
Pinang
Kalianyar
Pengayaan Dua
Pengayaga
Duri Kepa
Kayu Putih
Gondrong
Meruya Selatan
Jogja
Bunegara Kaum
Kunciran Indah
Alam
Pondok Betung
Rempoa
Lengkong Wetan
Jombang
Cilangkap Timur
Tanjung Barat
Jatikram
Setu

senseable city lab
BBVA
1000€ 500€ 100€
APRIL 2011 TRANSACTIONS IN
GROCERIES GAS STATIONS FASHION BAR
6h
12h
18h
24h


Automatic Fraud Discovery
AYASI
Fraudulent Transactions
Few
Many
igh concentration of fraud
re automatically characterized
r domain experts to take action.


07:07
JEU
VEN
SAM
DIM
LUN
MAR
MER

Big Data for Development?

What is Global Pulse?

An innovation initiative of the Secretary-General supporting the UN System in developing the

enabling environment,

analytical capabilities,

and organizational capacity

to harness real-time digital information for sustainable development and humanitarian action

HYPOTHESIS:

When people's needs change, they change how they use digital services, and these changes leave recognizable patterns in big data…in real time.

Mining for patterns of collective behavior

What people say

online news

social media

retail advertising

What people do

online search

mobile phone usage

financial transactions

Three Opportunities for Development

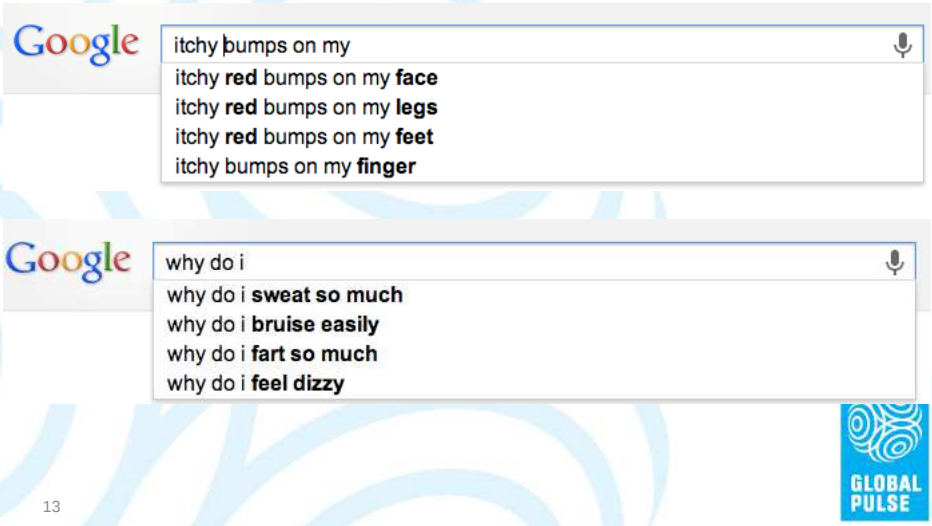
"Nowcasting"

Early Warning

Evaluation

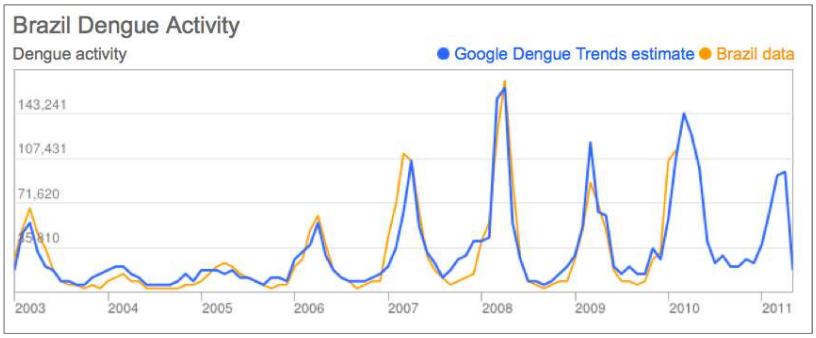
Services as Sensor Networks

Calling Dr. Google…



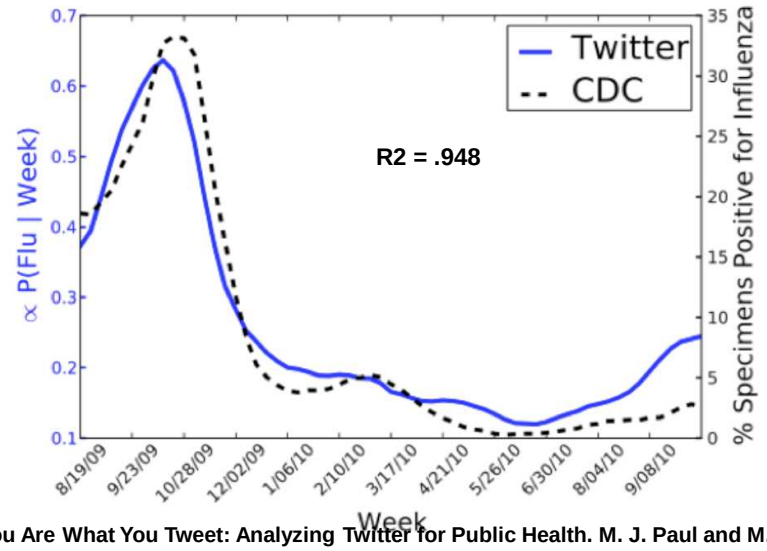
13

Google Dengue Trends



Volume of real-time searches for symptoms predicts predicts official # of cases of Dengue in Brazil

Twitter posts predict outbreaks…

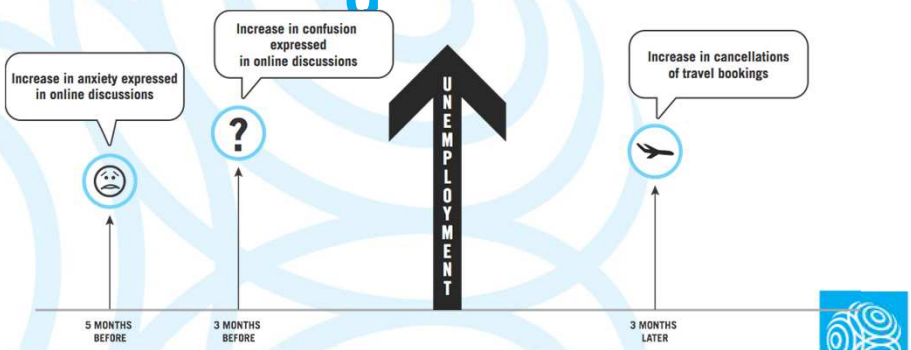


You Are What You Tweet: Analyzing Twitter for Public Health. M. J. Paul and M. Dredze, 2011.
http://www.cs.jhu.edu/%7Empaul/files/2011.icwsm.twitter_health.pdf

Can Social Media Predict Unemployment?

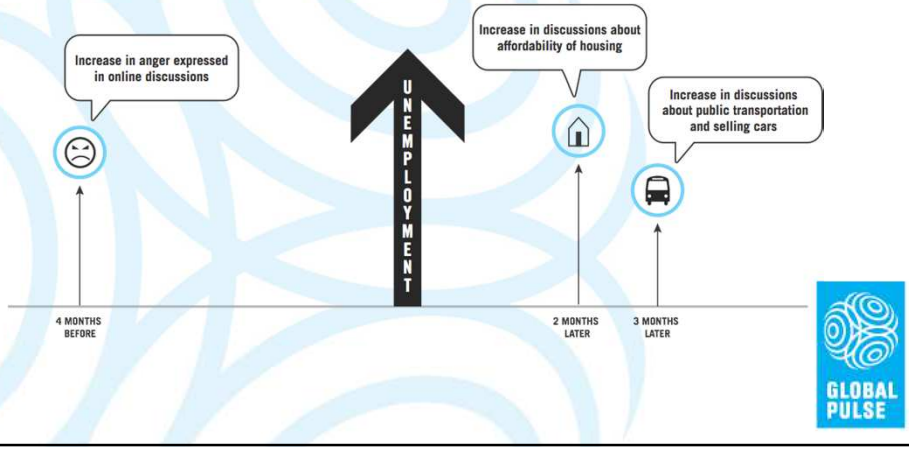
Online Discussions & Unemployment

Irelan d



Online Discussions & Unemployment

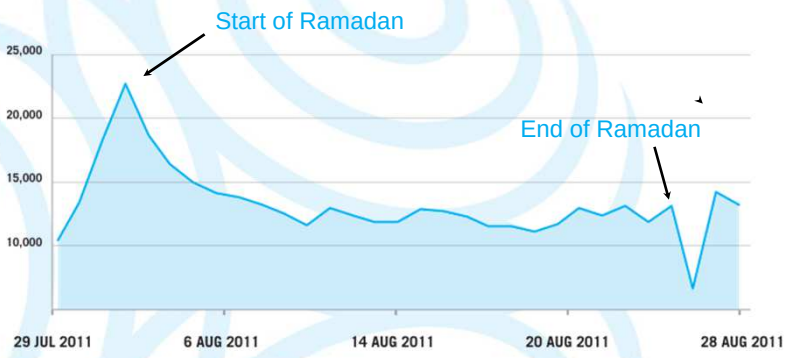
United States



Number of tweets per week related to food issues in Indonesia

GLOBAL PULSE

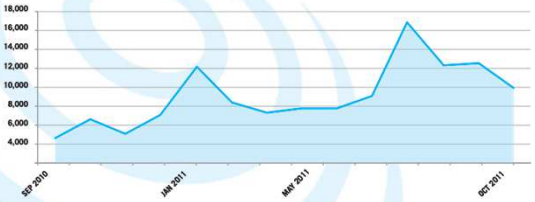
Tweets per day about food, during Ramadan in Indonesia



GLOBAL PULSE

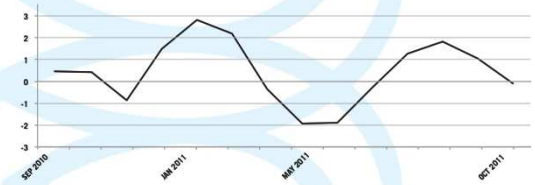
Tweets predict food basket inflation
(rice, chilies, fish, sugar, corn, cooking oil)

Tweets about the price of rice
(per month)



Official
Food Price Inflation

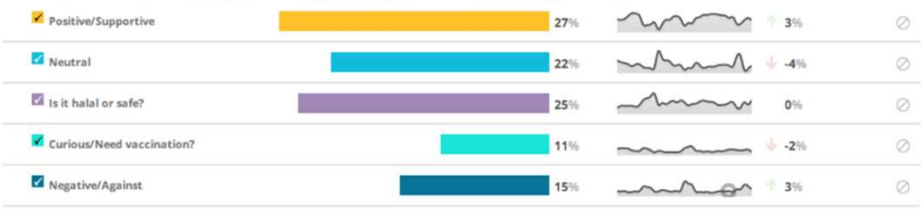
(monthly from 25 cities)



GLOBAL PULSE

Trending Twitter topic:
Are vaccines halal?

PULSE LAB JAKARTA

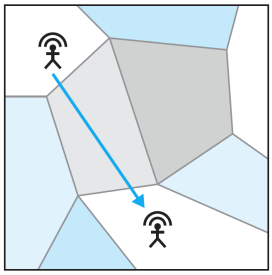


Cell Phone Network "Call Detail Records"

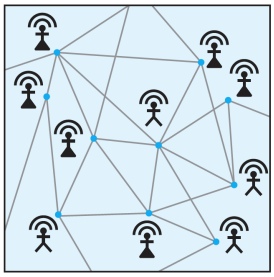
CDR Format:

CALLER ID	CALLER CELL TOWER LOCATION	RECIPIENT PHONE NUMBER	RECIPIENT CELL TOWER LOCATION	CALL TIME	CALL DURATION
X76VG588RLPQ	2°24' 22.14", 35°49' 56.54"	A81UTC93KK52	3°26' 30.47", 31°12' 18.01"	2013-11-07T15:15:00	01:12:02

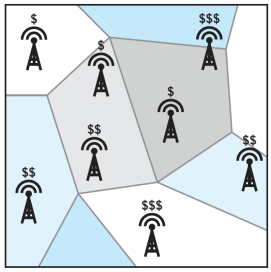
Mobility



Social Interaction

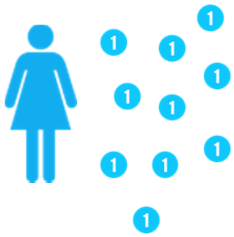


Economic Activity



Size and frequency of airtime purchases…

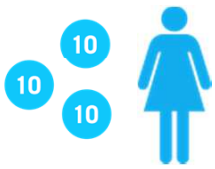
Subscriber A



Small top-offs every few days

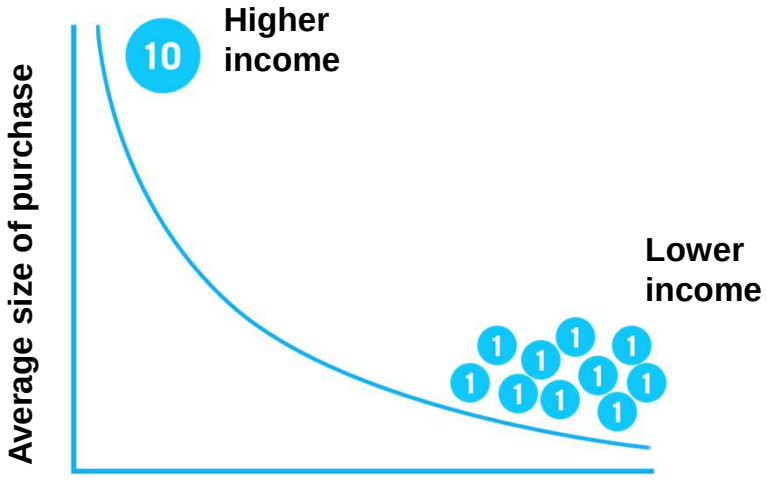
VS.

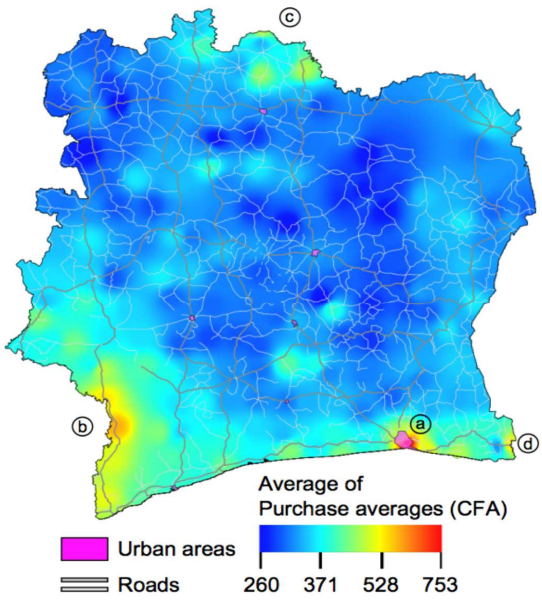
Subscriber B



Larger top-offs every few weeks

A proxy indicator of household income.



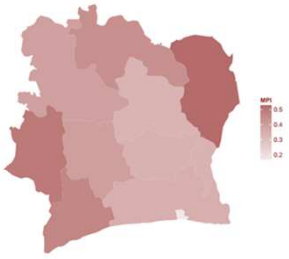


a) Abidjan
b) Liberian border
c) Roads to Mali and Burkina Faso
d) Road to Ghana

Ref: arxiv.org/abs/1309.4496: Evaluating Socio-Economic State Of A Country Analyzing Airtime Credit And Mobile Phone Datasets

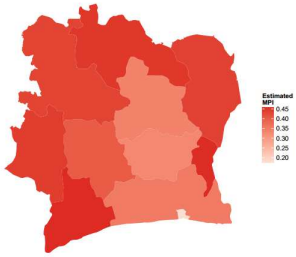
MAPPING SOCIOECONOMIC LEVELS OF POPULATIONS

Ground-Truth Data (Official)



Map of 11 regions showing Multidimensional Poverty Index

Using Mobile Phone Data

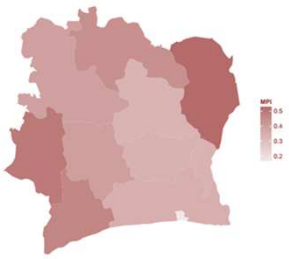


Poverty map estimated based on the link diversity antennas

Smith C, Mashadi A, Capra L (2013) Ubiquitous sensing for mapping poverty in developing countries, *Proceedings of the Third Conference on the Analysis of Mobile Phone Datasets.*

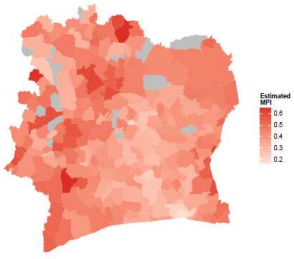
MAPPING SOCIOECONOMIC LEVELS OF POPULATIONS

Ground-Truth Data (Official)



Map of 11 regions showing Multidimensional Poverty Index

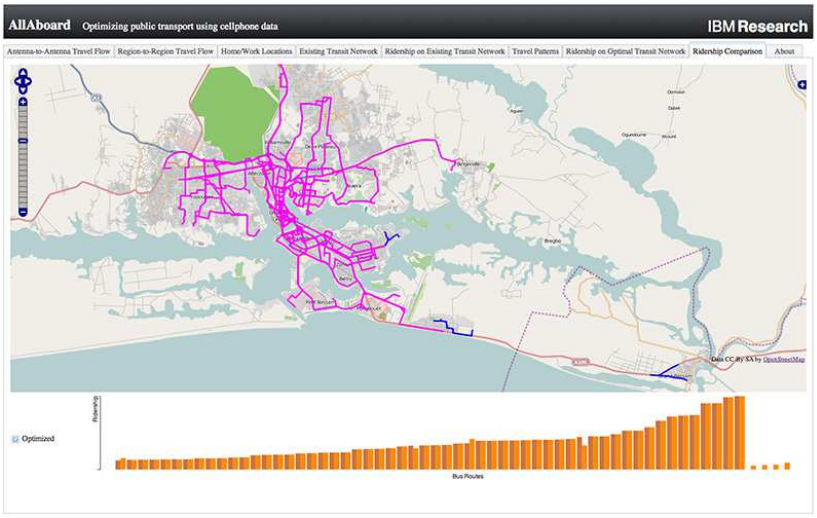
Using Mobile Phone Data



Finer granularity map estimated from communication patterns

Smith C, Mashadi A, Capra L (2013) Ubiquitous sensing for mapping poverty in developing countries, *Proceedings of the Third Conference on the Analysis of Mobile Phone Datasets.*

OPTIMIZING PUBLIC TRANSPORTATION NETWORKS



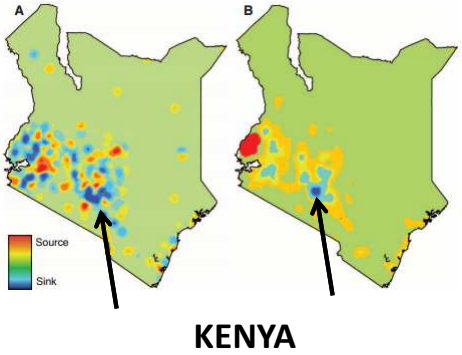
Berlingerio M, Calabrese F, Lorenzo G, Nair R, Pinelli F, Sbodio M (2013) AllAboard: A system for exploring urban mobility and optimizing public transport using cellphone data, *Proceedings of the Third Conference on the Analysis of Mobile Phone Datasets.*

Modeling the spread of disease

Quantifying the Impact of Human Mobility on Malaria

Amy Wesolowski,[1,2] Nathan Eagle,[3,4] Andrew J. Tatem,[5,6,7] David L. Smith,[6,8] Abdisalan M. Noor,[9,10] Robert W. Snow,[9,10] Caroline O. Buckee[4,11]*

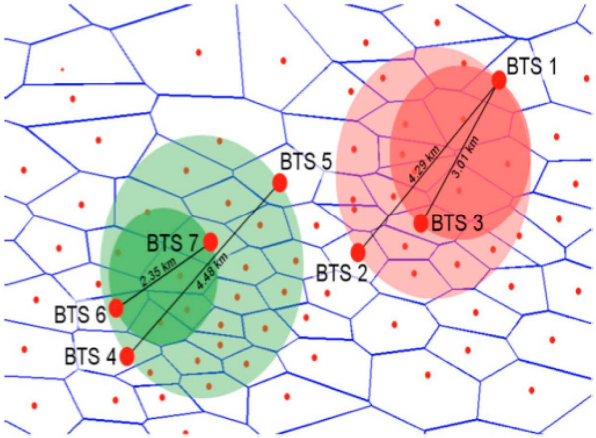
People movements Malaria parasites



KENYA

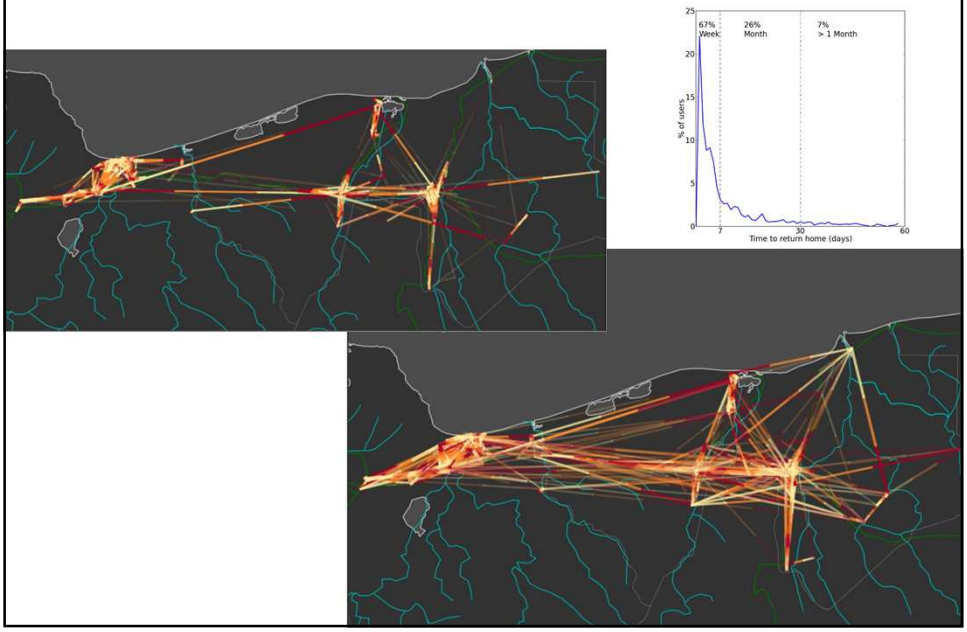
Wesolowski A, Eagle N , Tatem A, Smith D, Noor A, Snow R, Buckee C (2012) Quantifying the impact of human mobility on Malaria, *Science* 338(6104) 267-270.

Combating H1N1 in Mexico

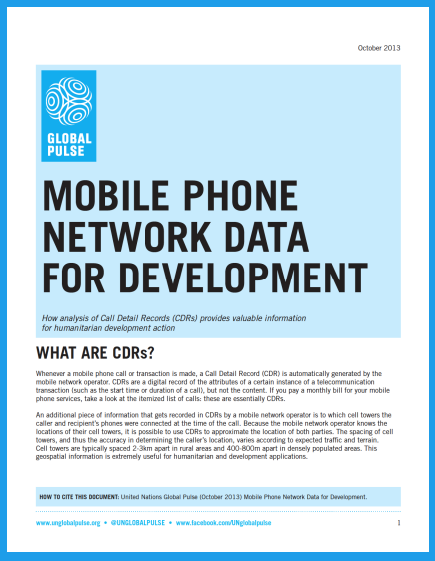


Enrique Frias-Martinez, Graham Williamson and Vanessa Frias-Martinez (2011) An Agent-Based Model Of Epidemic Spread Using Human Mobility and Social Network Information, *3rd International Conference on Social Computing, SocialCom '11, Boston.*

Quantifying Displacement during Natural Disasters



Our Latest Primer on Applications of Big Data



October 2013

GLOBAL PULSE

MOBILE PHONE NETWORK DATA FOR DEVELOPMENT

How analysis of Call Detail Records (CDRs) provides valuable information for humanitarian development action

WHAT ARE CDRs?

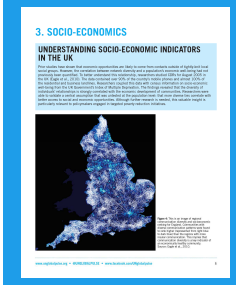
Whenever a mobile phone call or transaction is made, a Call Detail Record (CDR) is automatically generated by the mobile network operator. CDRs are a digital record of the attributes of a certain instance of a telecommunication transaction (such as the start time or duration of a call), but not the content. If you pay a monthly bill for your mobile phone services, take a look at the itemized list of calls: these are essentially CDRs.

An additional piece of information that gets recorded in CDRs by a mobile network operator is to which cell towers the caller and recipient's phones were connected at the time of the call. Because the mobile network operator knows the locations of their cell towers, it is possible to use CDRs to approximate the location of both parties. The spacing of cell towers, and thus the accuracy in determining the caller's location, varies according to expected traffic and terrain. Cell towers are typically spaced 2-3km apart in rural areas and 400-800m apart in densely populated areas. This geospatial information is extremely useful for humanitarian and development applications.

HOW TO CITE THIS DOCUMENT: United Nations Global Pulse (October 2013) Mobile Phone Network Data for Development.

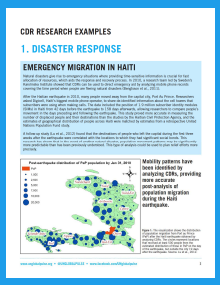
www.unglobalpulse.org • @UNGLOBALPULSE • www.facebook.com/UNglobalpulse 1

http://bit.ly/GPMobilePrimer



3. SOCIO-ECONOMICS

UNDERSTANDING SOCIO-ECONOMIC INDICATORS IN THE UK



CDR RESEARCH EXAMPLES

1. DISASTER RESPONSE

EMERGENCY MIGRATION IN HAITI

Mobility patterns have been identified by analyzing CDRs, providing more accurate estimates of population migration during the Haiti earthquake.

Big data represents an extraordinary opportunity to improve the lives of people everywhere…

…unless we fail to protect their privacy in the process.

Big data is also the greatest human rights issue of our generation.

Operating Principles

1. Radical transparency
2. Never handle personally identifiable information
3. Never handle confidential information
4. Never seek to re-identify anonymized data
5. Always assess privacy impact of research
6. Adhere to all applicable national and international data protection laws

GLOBAL PULSE

Reframing a polarized debate:

Big data is a *raw global good*

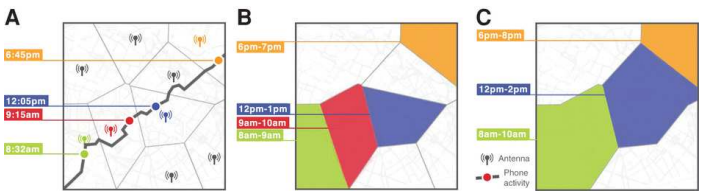
Towards a Risk-Based Approach

1. Privacy is a human right.
2. So is access to food, water, health, shelter, education…
3. You will be held accountable for how you use big data.
4. You will also be held accountable for failing to use it.
5. Re-identification risk is always non-zero

GLOBAL PULSE

We Need a Risk-Based Approach!

- Privacy is a human right.
- So is food, water, shelter, health, and education
- Risk of re-identification can never be zero
- Policymakers will be held accountable for how big data is used, **and for how it is not used.**

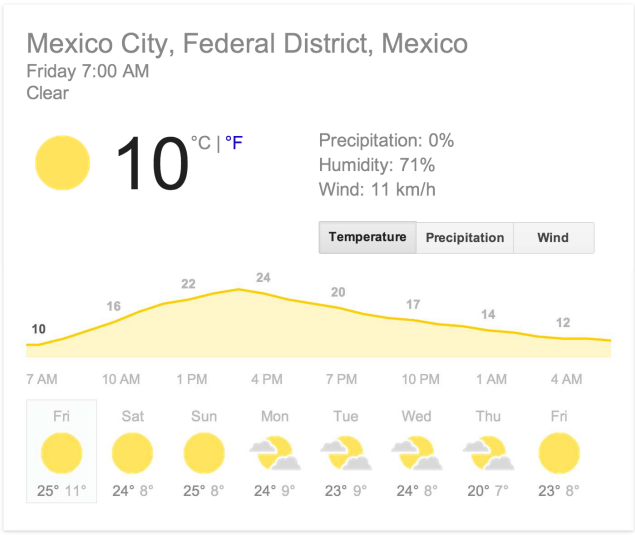


Source: http://www.nature.com/srep/2013/130325/srep01376/full/srep01376.html

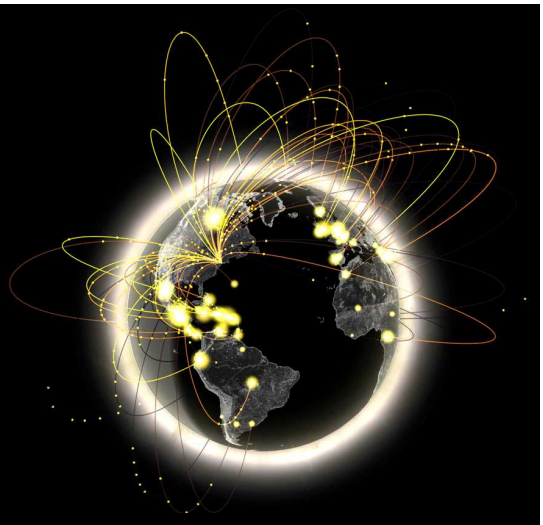
WAYS TO SHARE BIG DATA

- Provide pro-bono, or license-free, access to publicly available data (e.g., social media or news aggregators)
- Share aggregated and derived data sets for analysis under non-disclosure agreements (NDA)
- Allow researchers to analyze data within the private company's own network, under NDA
- Real-Time Data Commons: data pooled and aggregated between multiple companies of the same industry to protect competitiveness.
- Public/Private Alerting Network: companies mine data behind their own firewalls and share indicators.

A Real-time Public/Private Data Commons



ROBERT KIRKPATRICK
Director
UN Global Pulse
kirkpatrick@un.org
@rgkirkpatrick
+1 650-796-5709



THANK YOU!

Image credit: Aaron Koblin
24 hours of global phone calls and Internet traffic flowing through New York City