

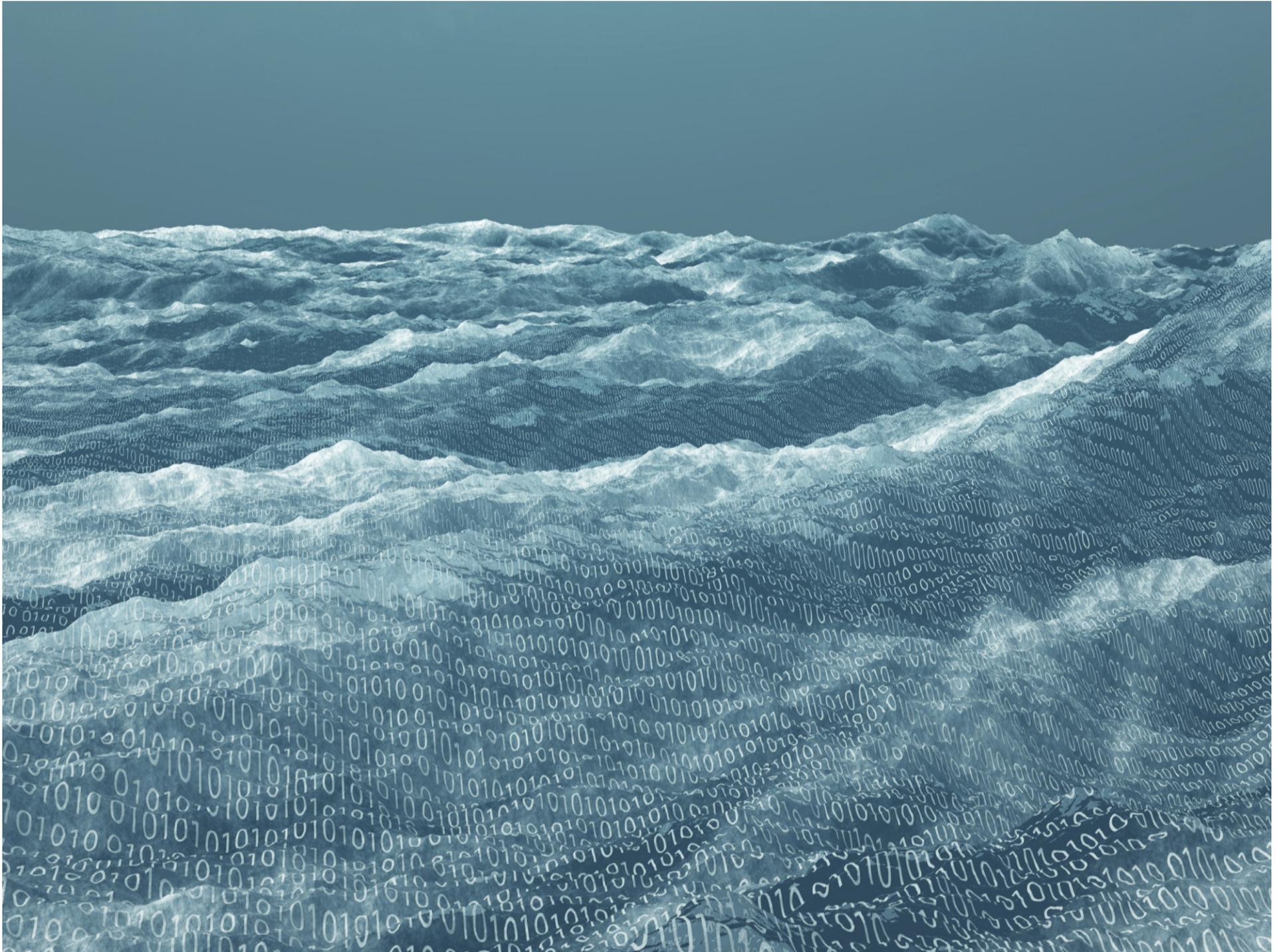


Driving a Big Data Revolution for Sustainable Development and Humanitarian Action

Robert Kirkpatrick
Director

www.unglobalpulse.org





Multidimensional recording of the flow of social dynamics...



survey data



big data

What if you could design continuous feedback into your programmes?

BIG (BEHAVIORAL) DATA

“What People Say”

Online news

Social media

Retail advertising

Radio & TV

“What People Do”

Online search

Mobile phone usage

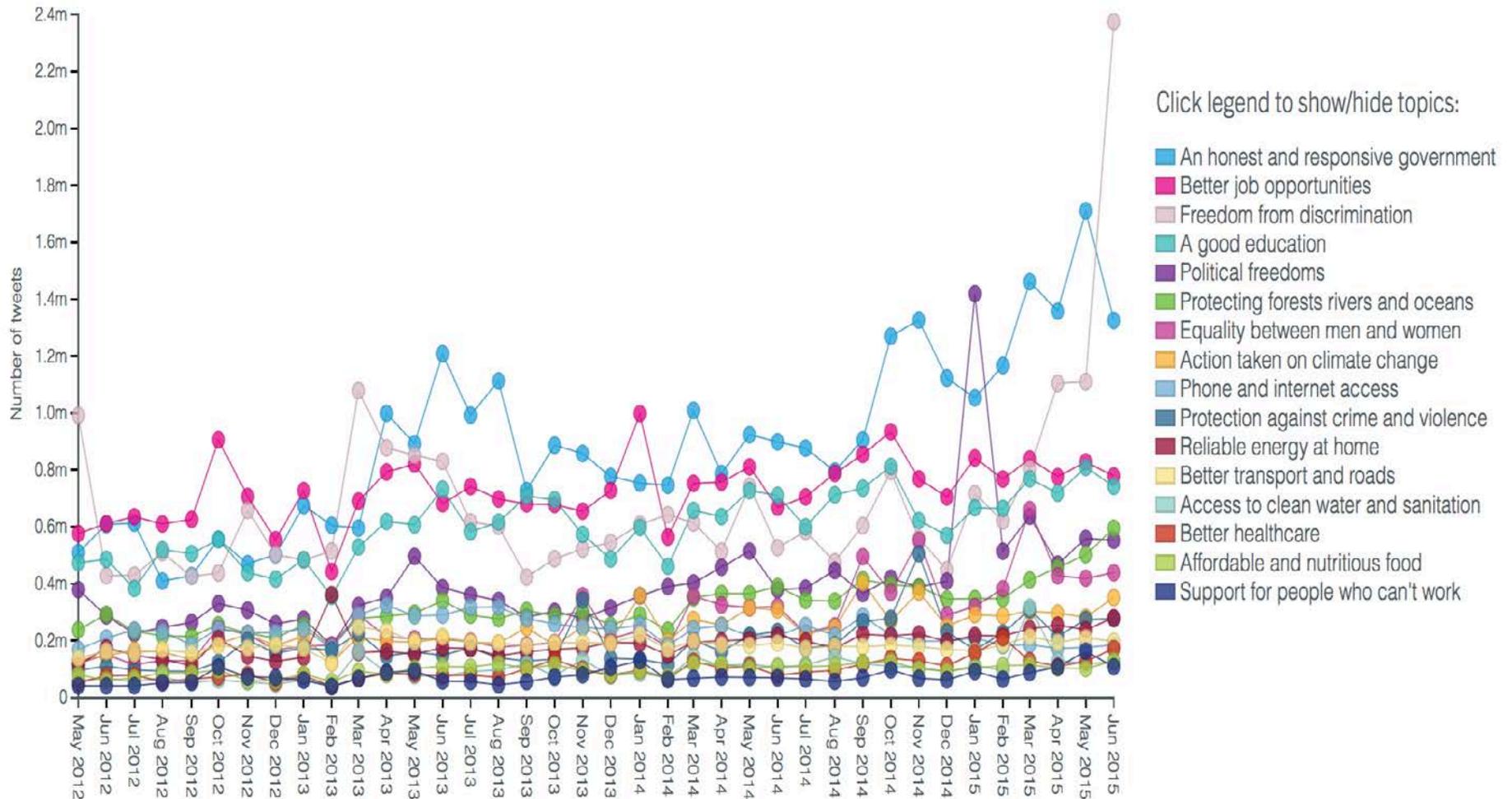
Transaction records

Postal traffic



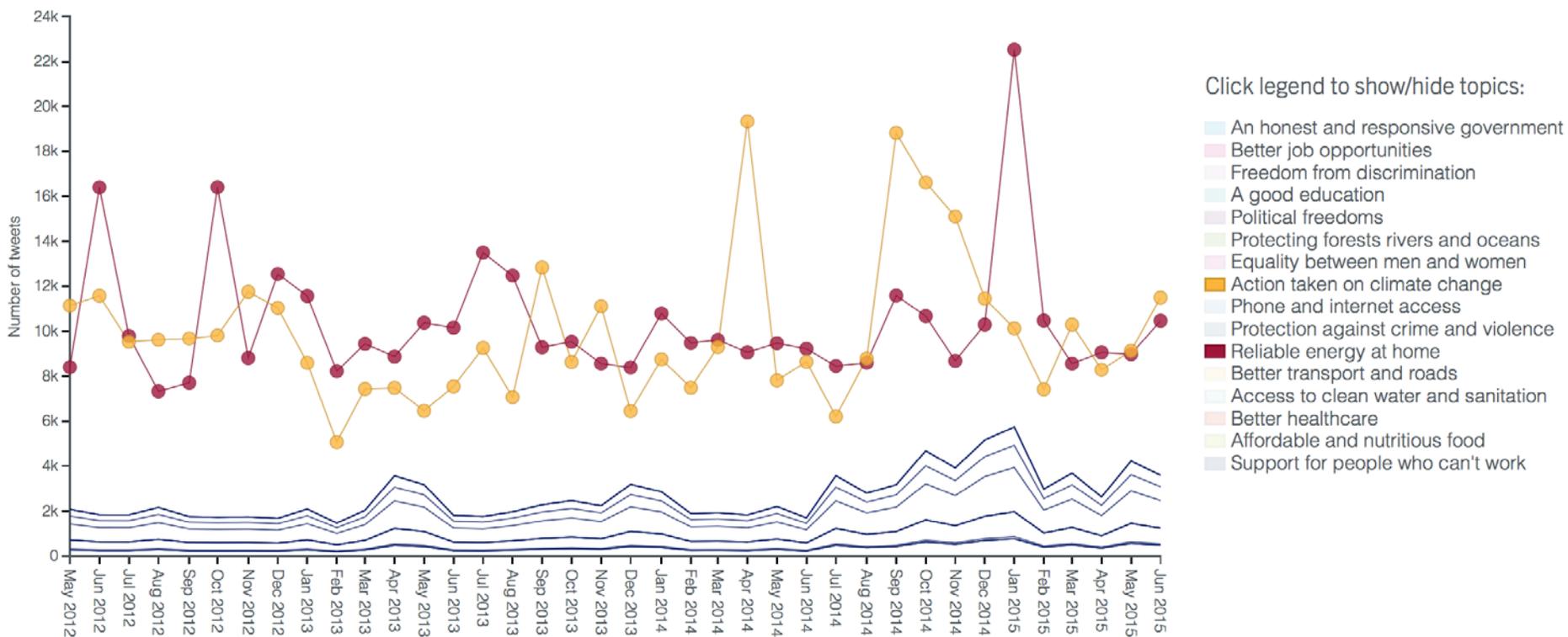
Trends: Number of tweets per month

- All Countries -



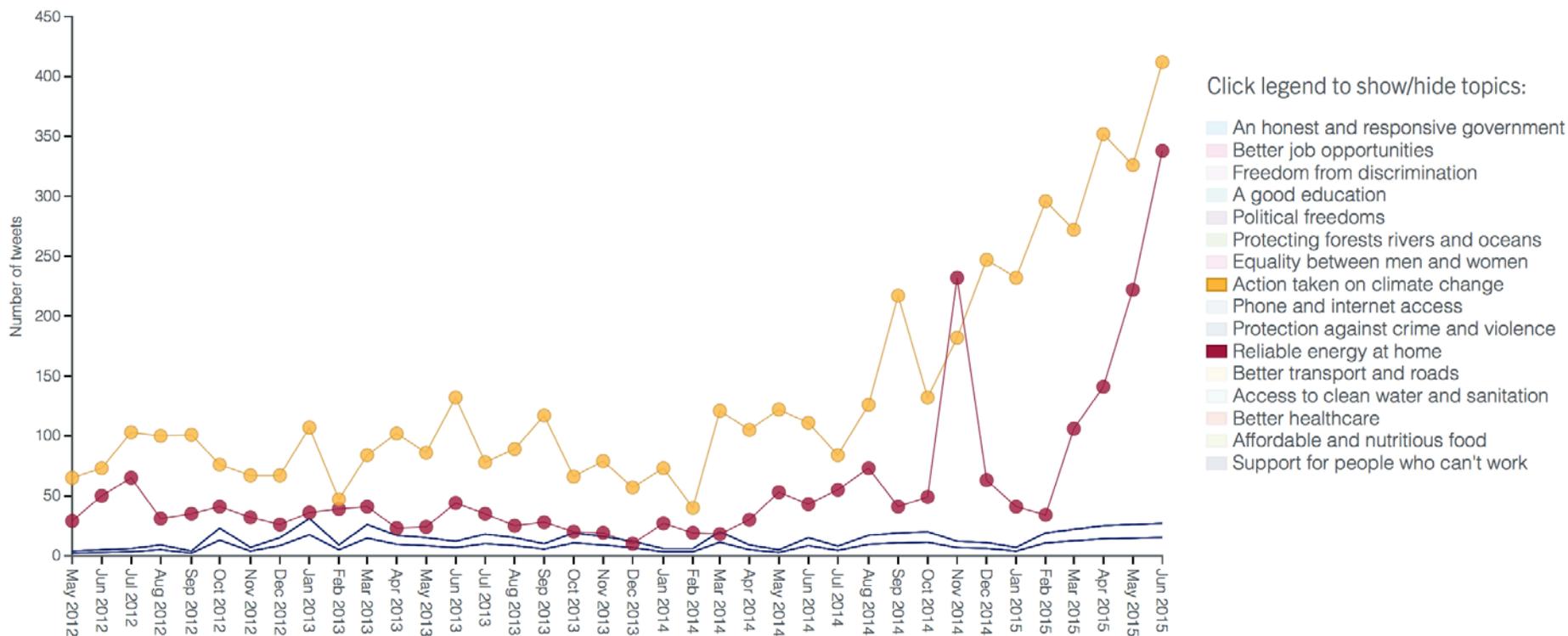
Trends: Number of tweets per month

Brazil



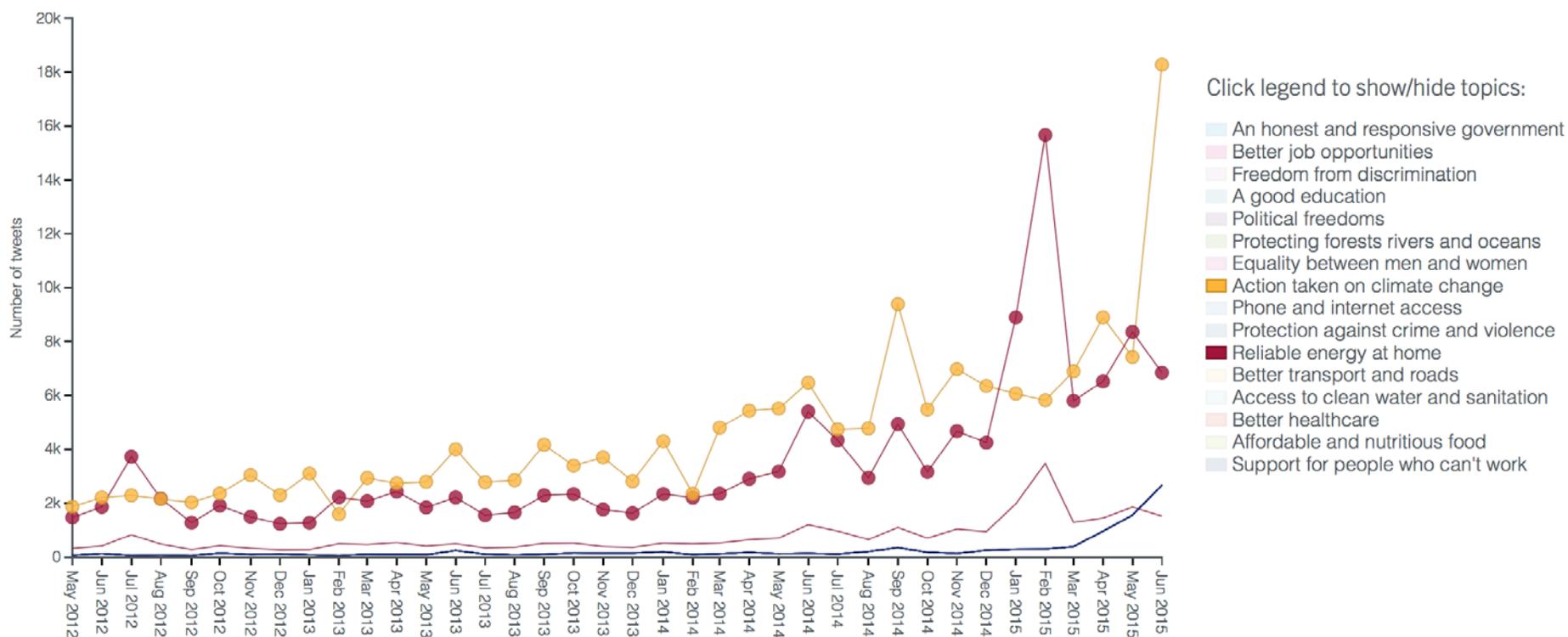
Trends: Number of tweets per month

Bangladesh



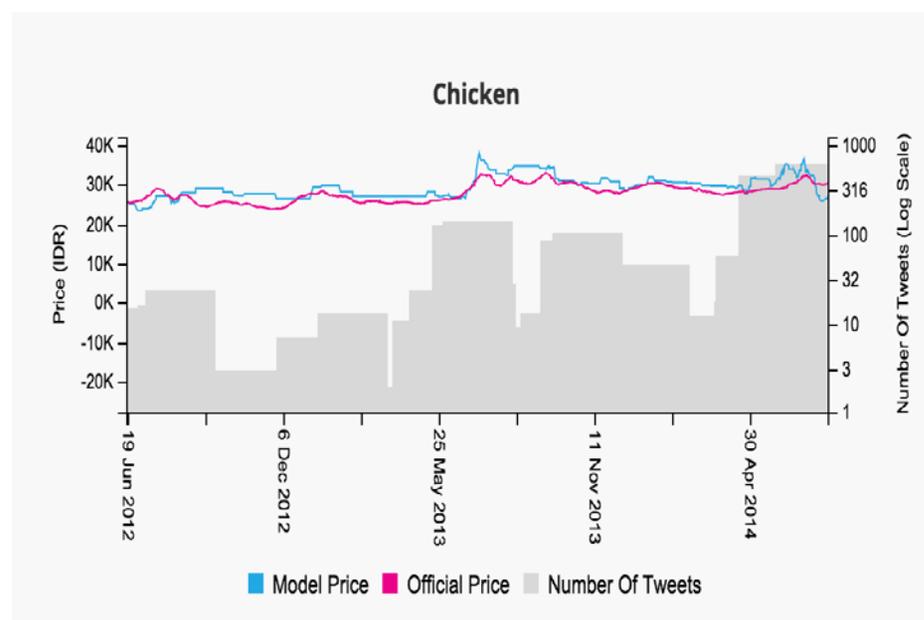
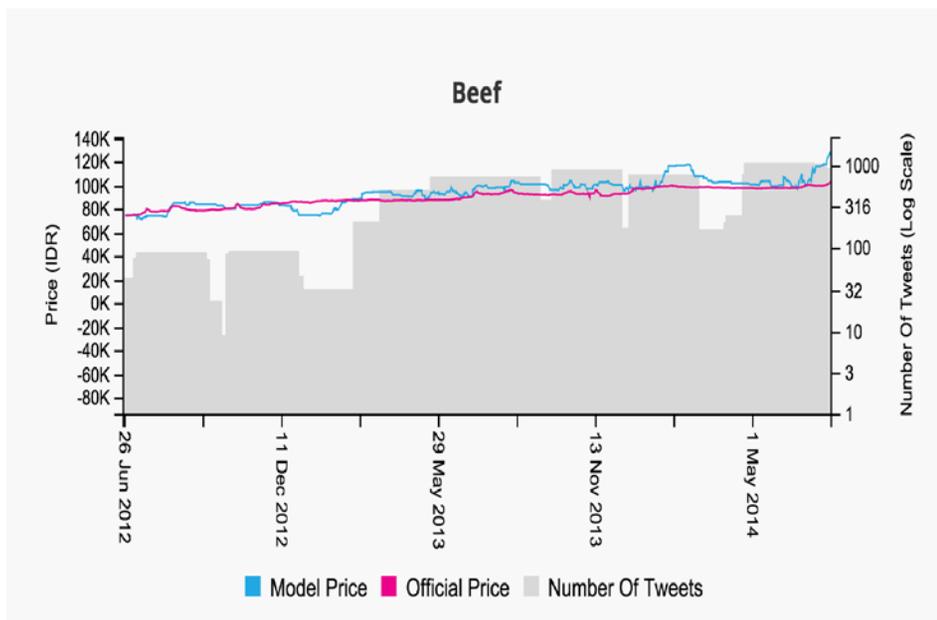
Trends: Number of tweets per month

India

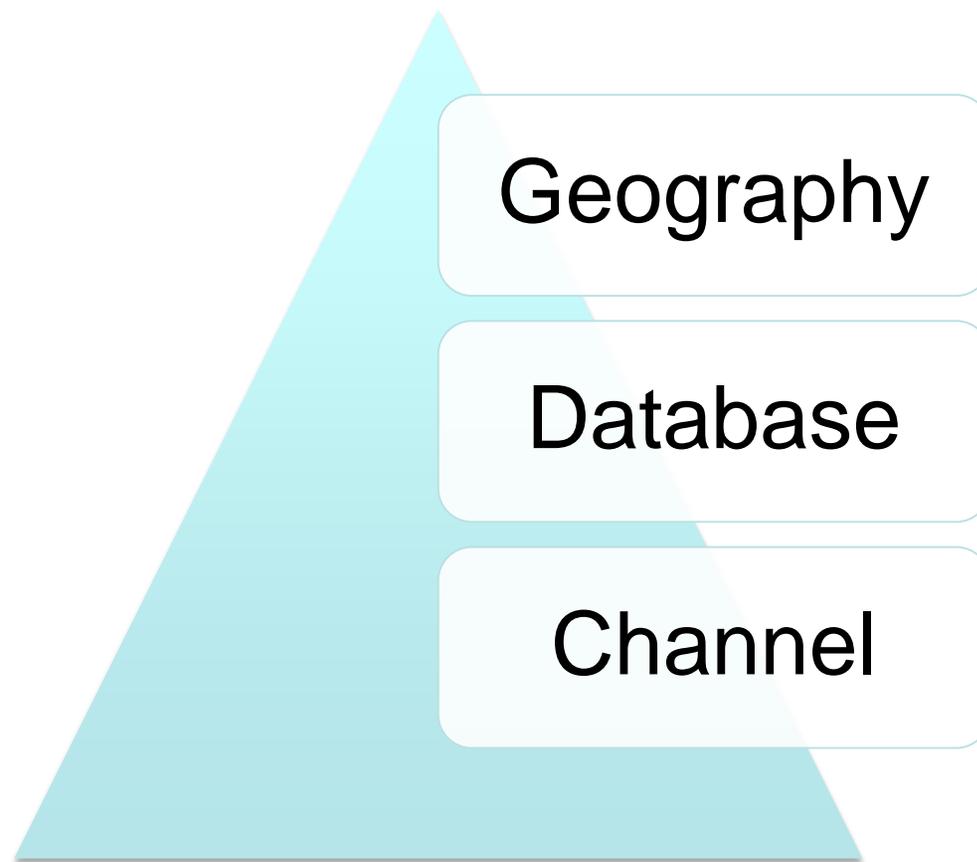




Nowcasting Food Prices

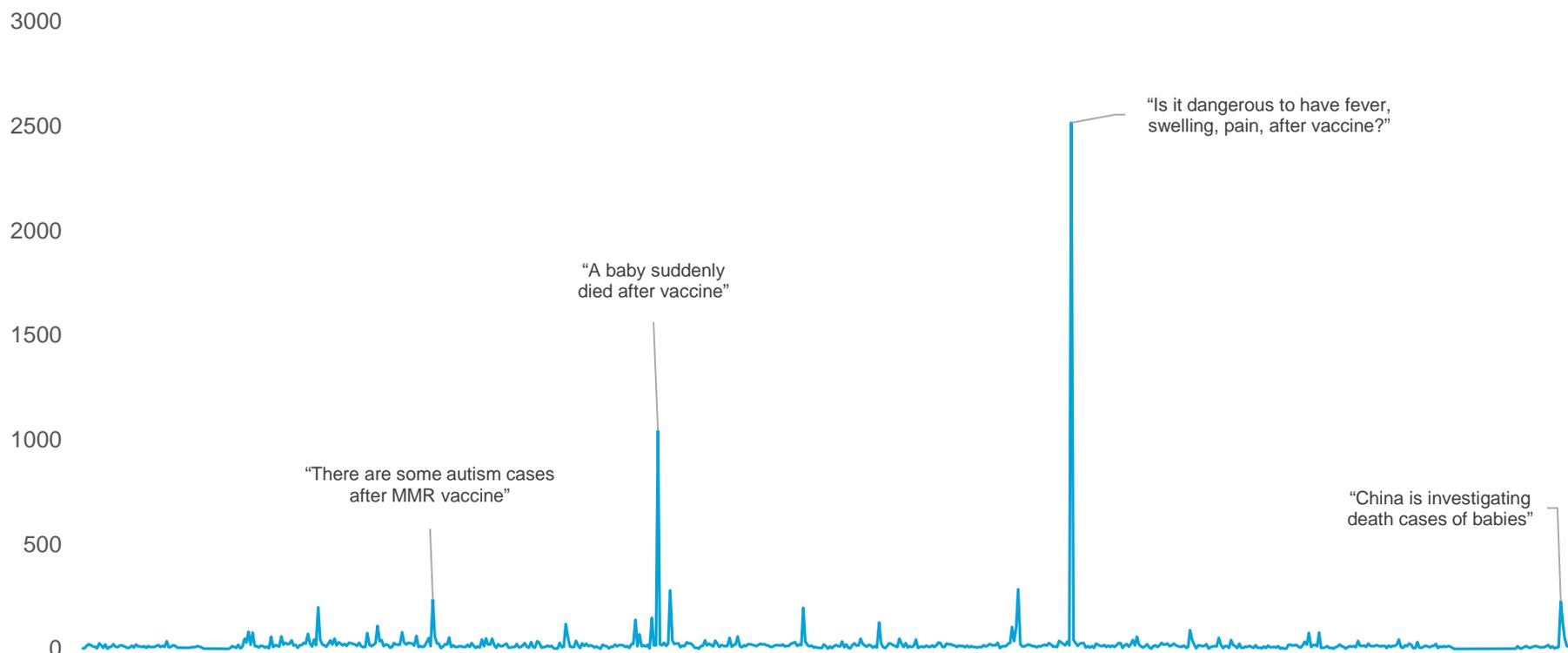


COMMODITY	PEARSON CORRELATION COEFFICIENT
Beef	0.87
Chicken	0.81
Onion	0.85





Immunization



Rank	2012-06-20	2012-10-08	2013-04-28	2013-12-23
1	Autism (213)	Death(1030)	Fever (1498)	Death (224)
2	Death (5)	Fever (14)	Swelling (1494)	Fever (3)
3	Sick (4)	Sick (4)	Pain (1491)	Crying (1)
4	Fever (2)	Crying (3)	Autism (1011)	Autism (1)
5	Crying (1)	Fever (3)	Fever (4)	-



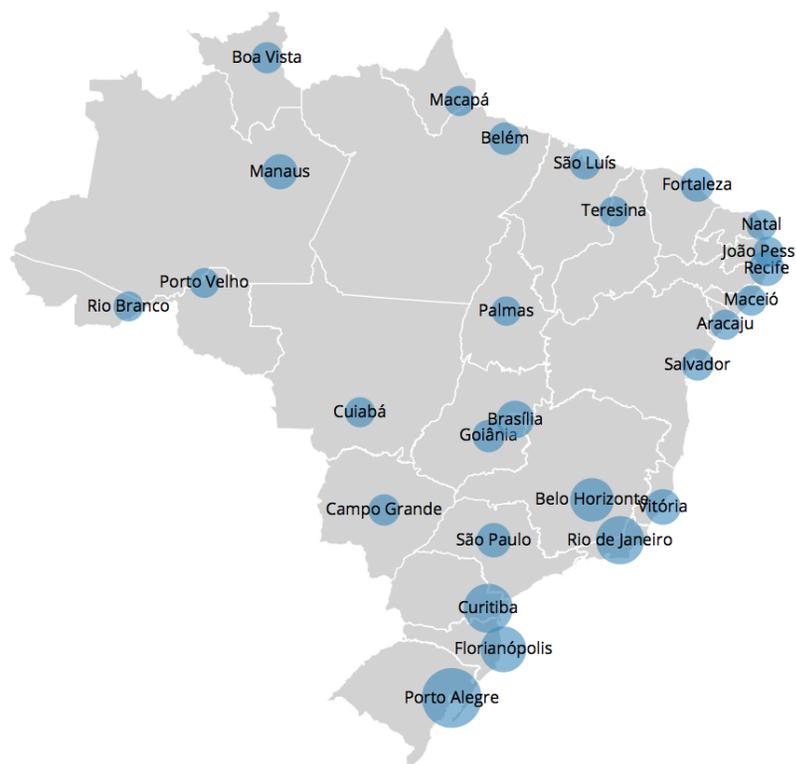
Project: HIV Risk Awareness at the World Cup

Partners: UNAIDS

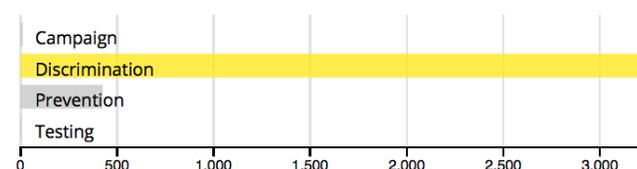
Tweets About Discrimination and HIV/AIDS

The interactive dashboard makes it possible to explore the volumes of tweets about HIV risk behavior generated each day during the 2014 World Cup. With simple mouse hover and clicks on charts, the data can be simultaneously viewed by multiple dimensions; time, host city, and the four selected categories, 'Discrimination', 'Prevention', 'Testing', and 'Campaign'.

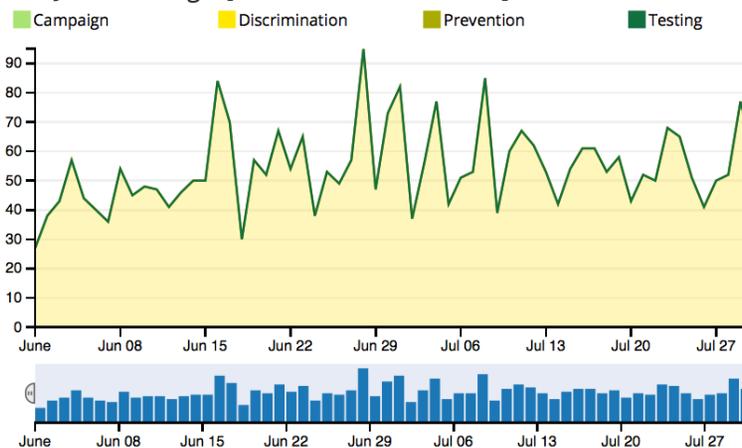
Brazil Map



Category reset



Daily Volume range: [06/01/2014 - 07/30/2014] reset

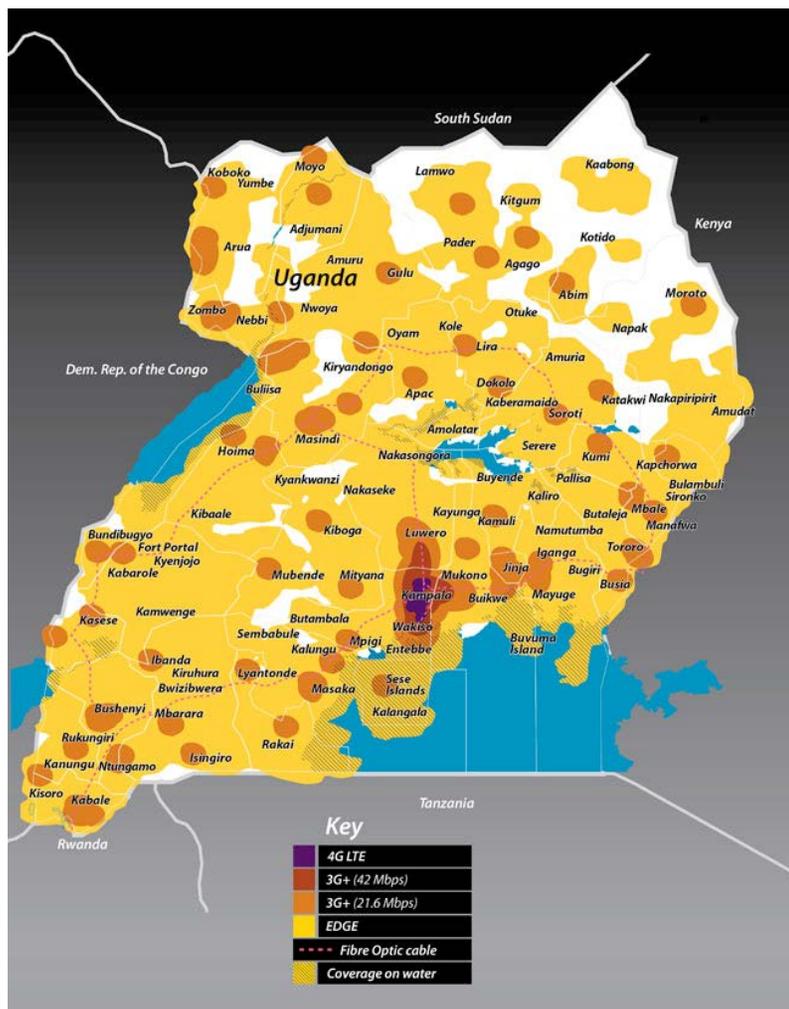


select a time range to zoom in

<http://www.unglobalpulse.org/social-media-HIV>



Real-Time Rural Talk Radio Analytics



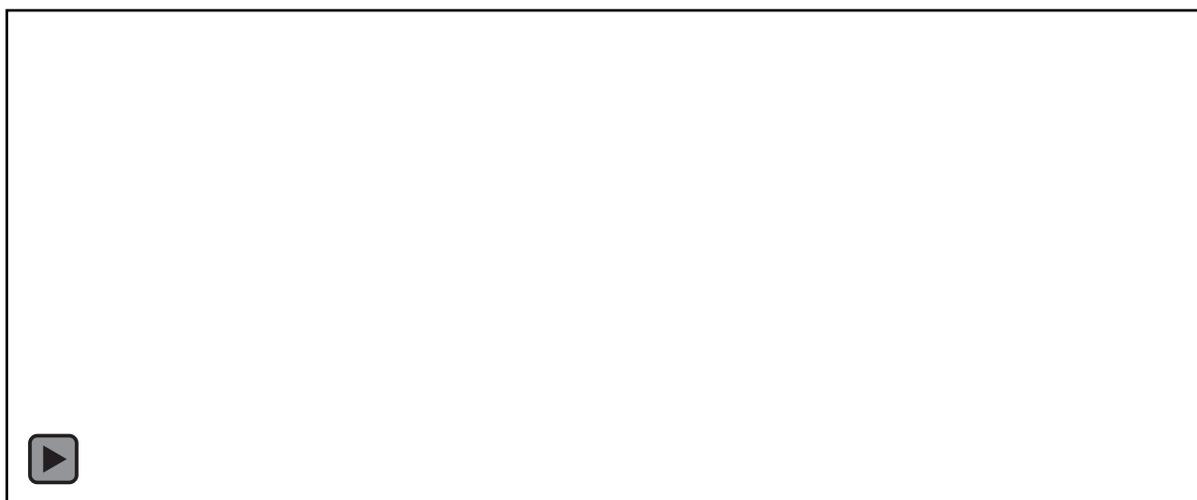
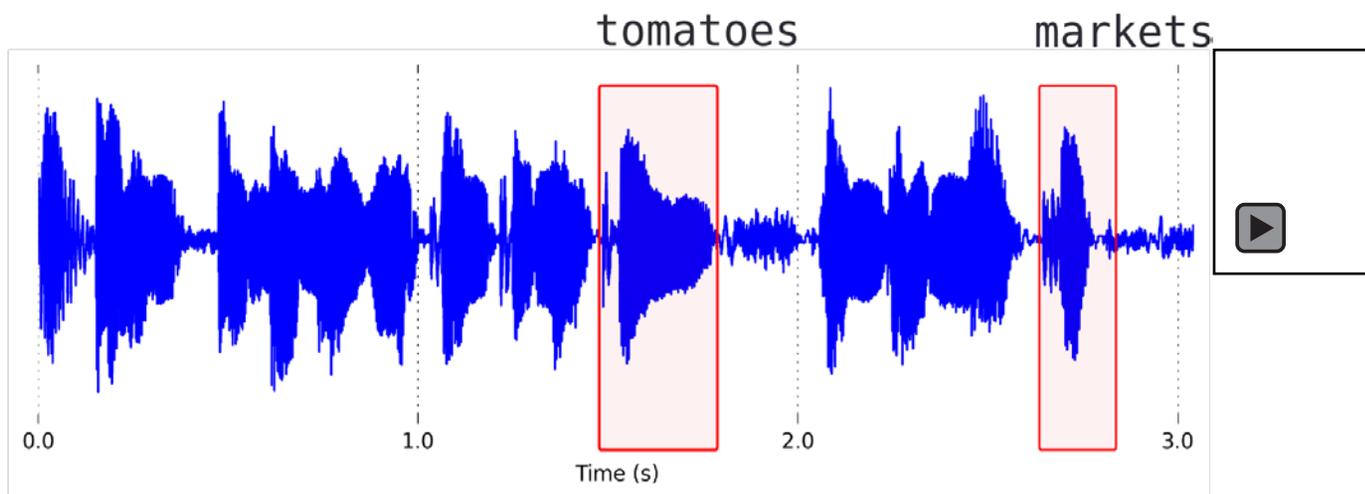
All 200 FM transmitters are in areas with mobile phone/EDGE reception, and most have 3G+ coverage.



Recording Ugandan talk radio on 50 stations 24x7 in English, Luganda and Acholi.



"you find so many rotten tomatoes in the markets"

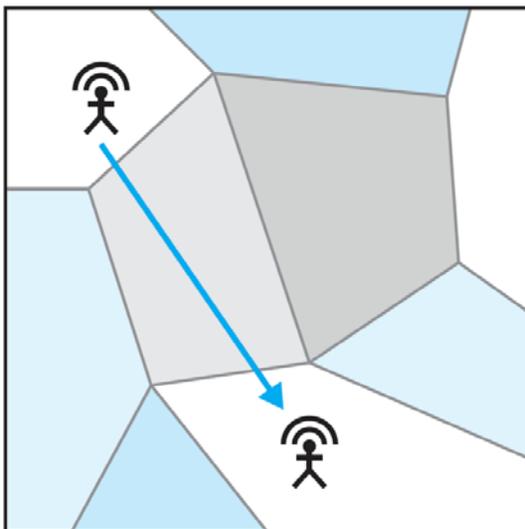


Call Detail Records

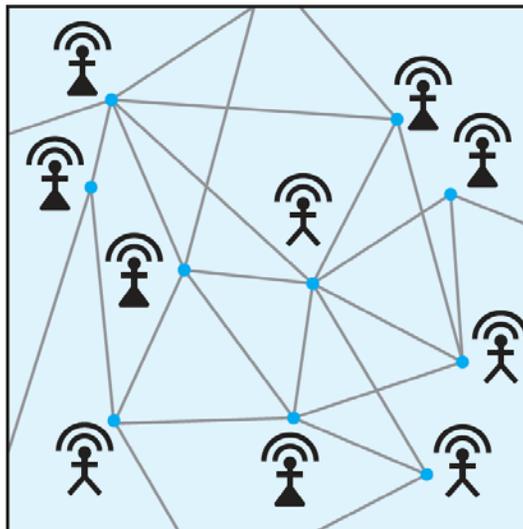
CDR Format:

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

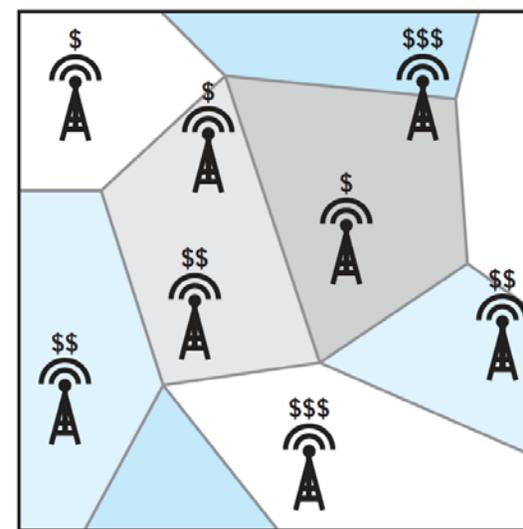
Mobility



Social Interaction



Economic Activity



HUMAN BEHAVIOURS IN FLOODS USING MOBILE PHONE DATA

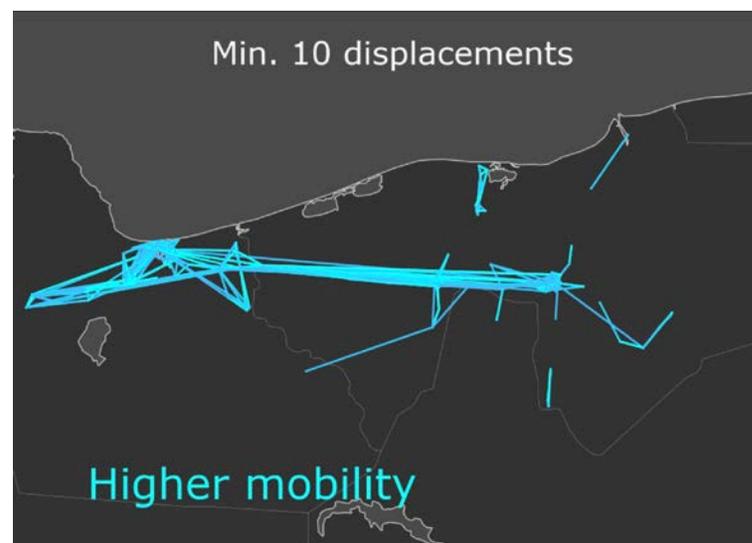


POPULATION DISPLACEMENT

TYPICAL NIGHT IN
TABASCO



NIGHT OF FLOODS:



PARTNERS: Global Pulse, WFP, Technical University of Madrid, Telefonica Research

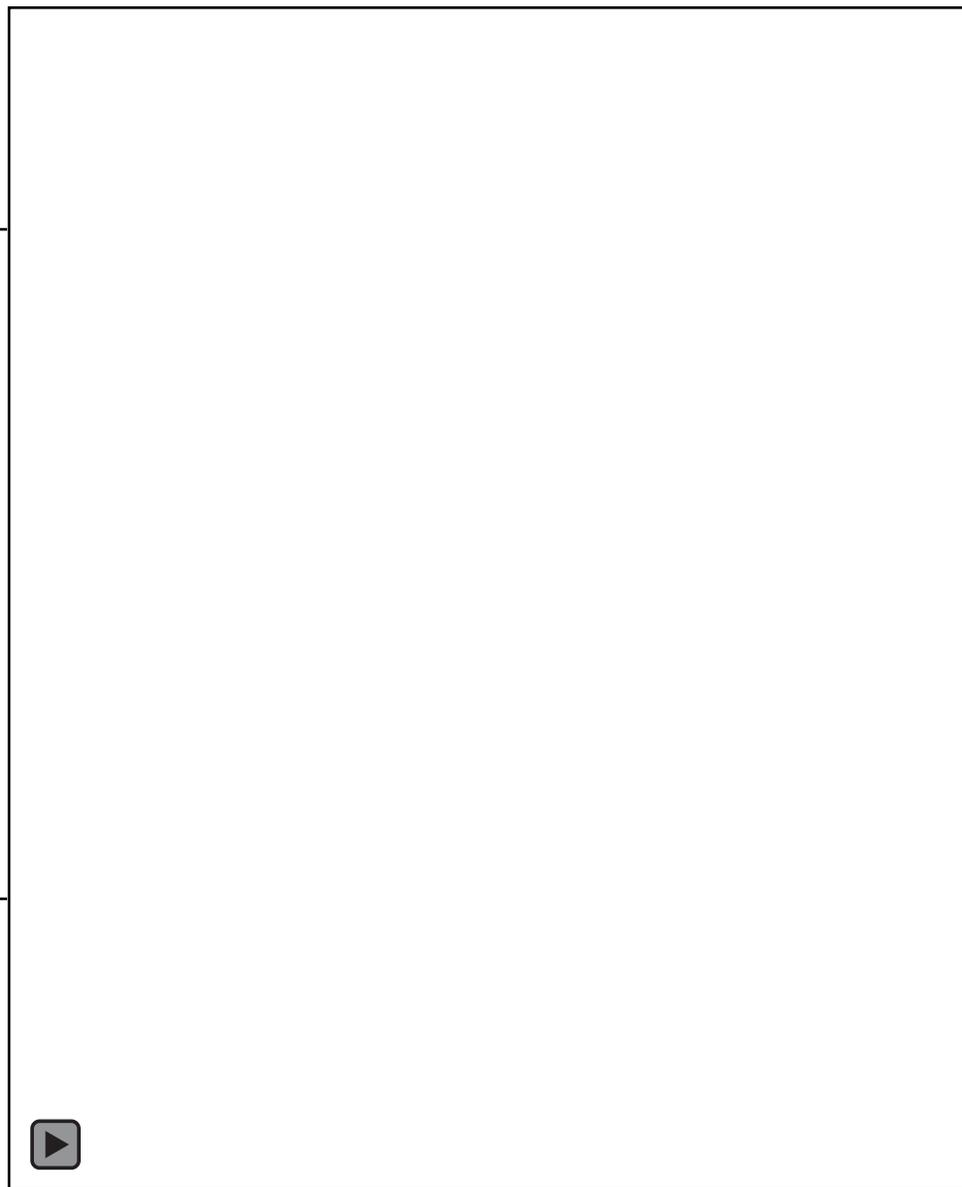
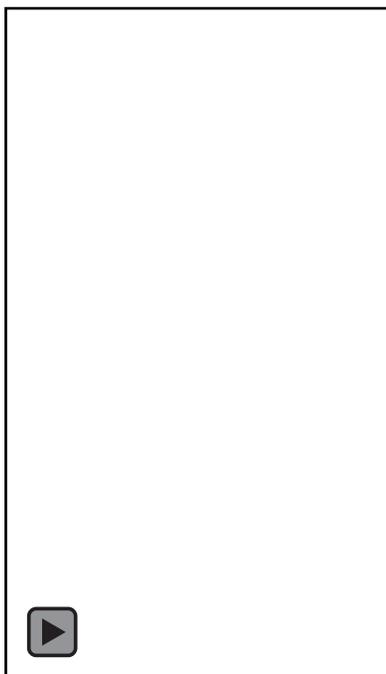


The University Of Sheffield.



World Health Organization

Predictive Epidemiology



It turns out that how much people spend money on mobile airtime credit...



...predicts how much they spend money on food.



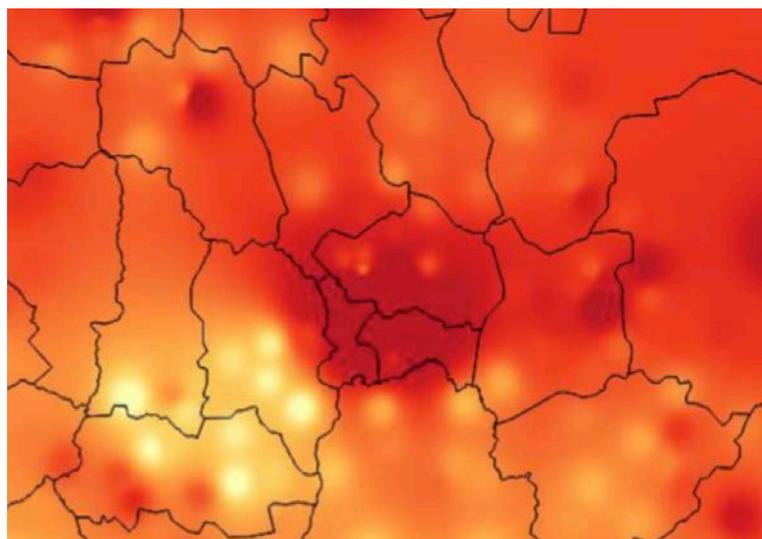
MOBILE AIRTIME PURCHASES PREDICT FOOD CONSUMPTION



World Food Programme

We can generate a real-time estimate of:

- Household food consumption, 89% accuracy
- Multidimensional Poverty Index, 80% accuracy



REAL-TIME FOOD SECURITY MAP

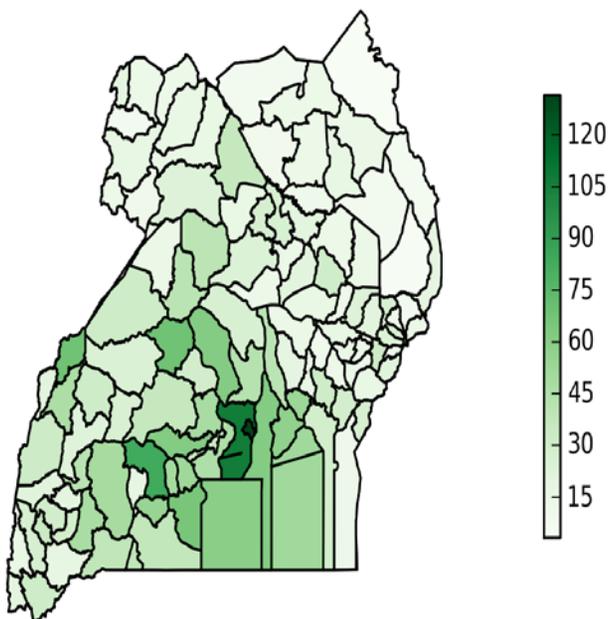
PARTNERS: WFP, UNIVERSITÉ CATHOLIQUE DE LOUVAIN, REAL IMPACT ANALYTICS



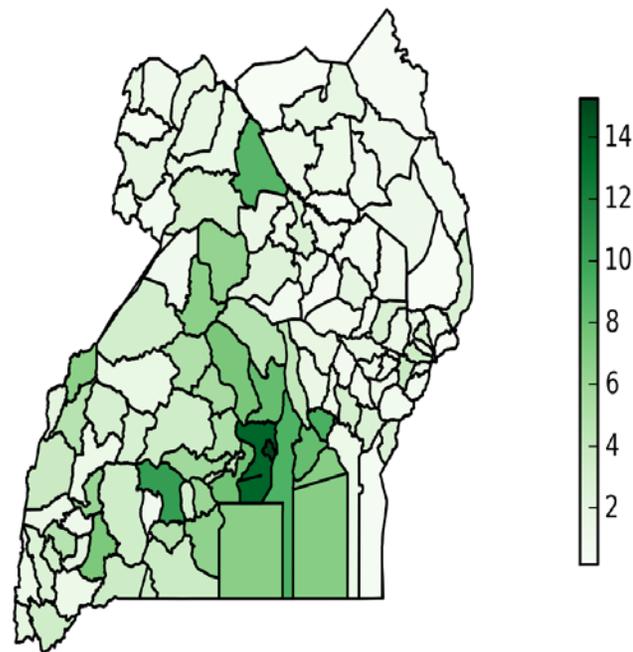
Real-Time Household Consumption Monitor



Mean monthly household consumption per adult equivalent ('000 UGX)



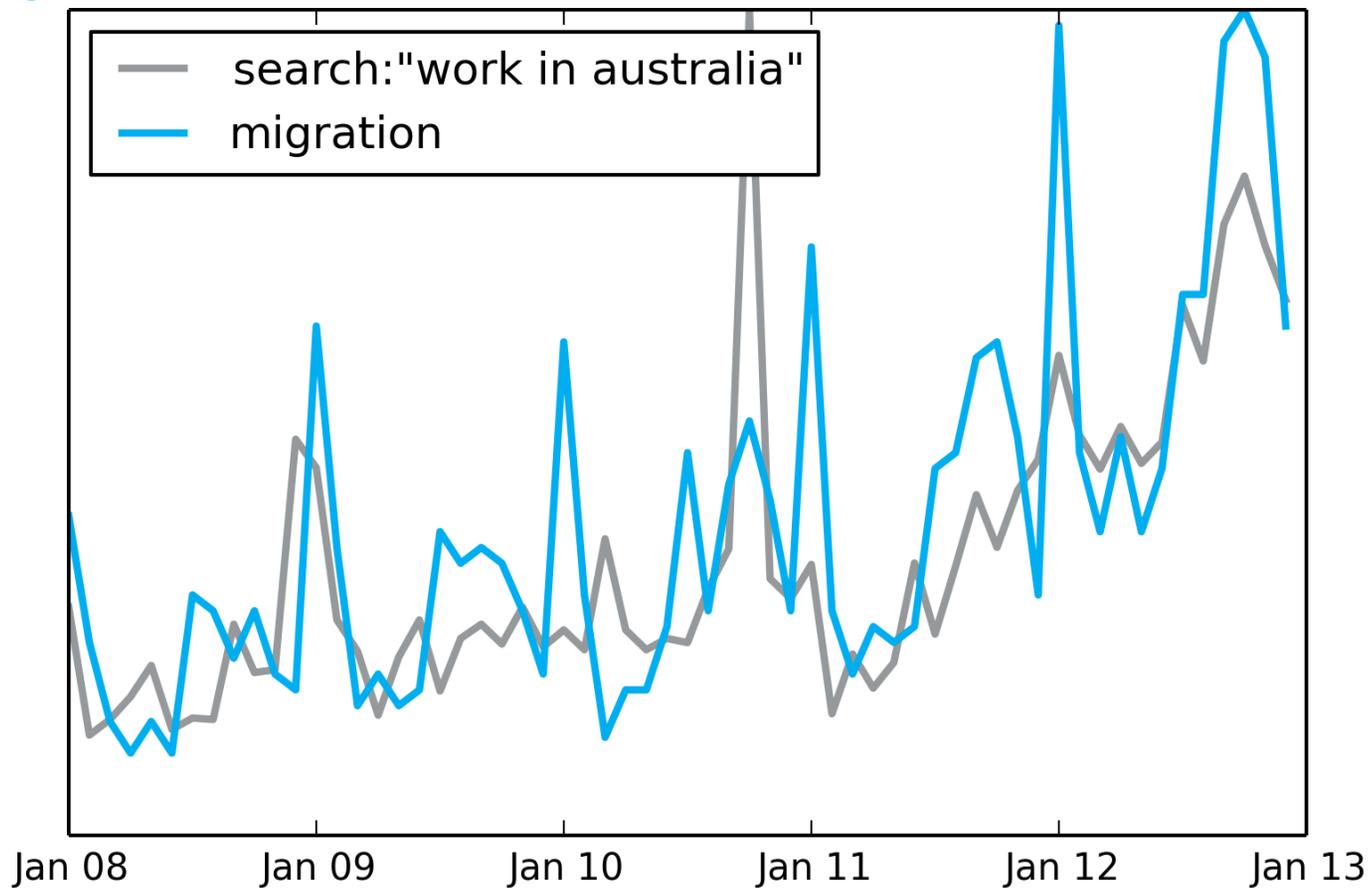
Mean monthly airtime expenditure ('000 UGX)



Airtime consumption can provide a real-time proxy measure of household consumption



USING ONLINE SEARCH DATA AS A PROXY INDICATOR OF INTERNATIONAL MIGRATION



Project: Exploring Aggregated Postal Flow Data for Insights on Economic Indicators

Partners: Universal Postal Union (UPU)



Challenges to Mainstream Analytics

- ## Adoption
1. **Few examples of successful implementation at scale**
 - a. Little data being shared by owners
 - b. Few technology solutions available
 - c. Limited uptake by humanitarian and development organizations
 - d. Low public sector investment

 1. **Potential users facing systemic barriers:**
 - a. No data standards or best practices for sharing big data
 - b. Few standards for anonymization of big data
 - c. Regulatory environment does not adequately support balancing risk of misuse of data against risk of not using it for public good

 1. **Ecosystem needed for scaling up is embryonic**
 - a. Limited institutional capacity to integrate big data into operations
 - b. Data scientists and big data technologists aren't talking to development or humanitarian practitioners
 - c. Little cross-sectoral visibility of current activities
 - d. No common global agenda for big data innovation



Theory of Change

1. A critical mass of high-impact scaled implementations drives **public consensus** on right to have one's data used in one's interest and urgent need to address known risks.
 2. Private sector recognition that data sharing generates shared value enables **sustainable access** to relevant sources of historical and streaming data.
 3. Recognition of need to balance the risk of misuse against the risk of not using big data leads to standards and frameworks for **effective governance** of the use of big data for the public good.
1. Increased user demand and common understanding of priority opportunities drives emergence of both a thriving **community of practice** and a **vibrant market for solutions**, in turn resulting in **accelerated innovation**.
 2. By 2020, increased organizational buy-in and institutional capacity spurs a global shift to mainstream **adoption of data analytics** for measuring and achieving sustainable development.



Talkin' bout a revolution

- Big data will contribute to a **measurement revolution** through better, faster, less-costly tracking of specific SDG indicators
- Big data will contribute to an **accountability revolution** by virtue of the fact that government plays no role in either its collection or the timing availability of insights derived from it.
- Big data will contribute to the **management revolution** needed to achieve the SDGs in the first place, by enabling accelerated learning and adoption of agile, efficient, adaptive processes.



Assumptions underlying indicators

Hypothetical Indicator....

“% population living within 5 km of major roadway”

...vs. what we do several times per week:



Hypothetical Application: Real-Time Poverty Monitor for SDG 1

- A free and open source data analytics platform running inside the data centres of 100+ mobile operators in LDCs, generating information accessible to users via web browser and a mobile app.
- Analyzes mobile airtime purchase records to continuously generate an 80% accurate estimate of the MPI.
- The model is augmented by integration of other information including both statistical survey data as well as data analytics approaches such as:
 - Satellite-based analysis of nightly light emissions
 - Automated recognition of roofing materials of new structures
 - Trade estimation via real-time postal traffic analysis
 - Trends in retail purchasing behavior.



Real-Time Poverty Monitor, cont.

TARGET 1.1: By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than \$1.25 a day.

Measurement.

View heat map of poverty estimates down to the sub-district level. View areas where recent spending is significantly above or below anticipated levels. Click on sub-district to view more detailed map and changes over the past 3 years.

Achievement. Automated alerts and notifications triggered based on anomalous drops in spending levels, particularly in at-risk populations. These may be investigated and verified via mobile surveys, and used to target measures such as cash transfer programmes, school feeding programmes, and accelerated access to insurance for farmers.

Nota bene: efficiency gains resulting from using data already being passively produced will free up resources for implementation. But analytics are not free. ROI will need to be determined empirically.



Real-Time Poverty Monitor, cont.

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

Measurement. Disaggregation by gender and age group based on patterns of subscribers' social behavior (locations where calls are made, call duration, size of social network, responsiveness to texts, etc.).

TARGET 1.3: Implement nationally appropriate social protection systems and measures for all, including floors, and by 2030 achieve substantial coverage of the poor and the vulnerable.

Achievement. Real-time media mining (analysis of news, social media, radio and local television) for perceptions of social protection services, reports of difficulties qualifying, grievances and redress issues could all be used to respond proactively to emerging risks to implementation.



Real-Time Poverty Monitor, cont.

TARGET 1.4: By 2030, ensure that all men and women, in particular the poor and the vulnerable, have equal rights to economic resources, as well as access to basic services, ownership and control over land and other forms of property, inheritance, natural resources, appropriate new technology and financial services, including micro finance.

Measurement. Analysis of mobile financial services data could be used to assess availability and uptake of mobile money transfers, access to mobile loans, and repayment default rates.

Achievement. Google search query analysis and media mining could be used to identify areas or demographics with unmet need, cases where a lack of information is hindering uptake, or difficulties qualifying. (Target 1.4)



Real-Time Poverty Monitor, cont.

TARGET 1.5: By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental shocks and disasters.

Measurement. Longitudinal analysis of the impacts of commodity price fluctuations and natural disasters on different communities' spending patterns, and of those communities' capacity to tap into financial support via mobile remittances, may be used to compute vulnerability to exogenous shocks.

Achievement. Media mining could be used to assess awareness of risk, response protocols, and perceptions of the effectiveness of responses in an ongoing fashion, both to target programme activities and assess effectiveness of resilience-building initiatives; mobility patterns derived from CDRs or other data sources could be used to assess compliance with evacuation routes during exercises or actual emergencies.



DATA FOR CLIMATE ACTION



Reflections on Question 3

- Representativeness: indicator, trend, or event?
- Online behavioral indicators may be cultural
- Online behavior evolves over time
- How to catch up with the data revolution
- Data Philanthropy vs. Sustainable access
- “We are what we habitually do”
- Governance: Privacy vs. Utility
- Accountability: open data vs. open insights

