UN Global Pulse

Table of Contents

GLOBAL PULSE

Global Pulse is an innovation initiative launched by the Executive Office of the United Nations Secretary-General, in response to the need for more timely information to track and monitor the impacts of global and local socio-economic crises. The Global Pulse initiative is exploring how new, digital data sources and real-time analytics technologies can help policymakers understand human well-being and emerging vulnerabilities in real-time, in order to better protect populations from shocks.

The initiative was established based on a recognition that digital data offers the opportunity to gain a better understanding of changes in human well-being, and to get real-time feedback on how well policy responses are working. The overarching objective of Global Pulse is to mainstream the use of data mining and real-time data analytics into development organizations and communities of practice. To this end, Global Pulse is working to promote awareness of the opportunities Big Data presents for relief and development, forge public-private data sharing partnerships, generate high-impact analytical tools and approaches through its network of Pulse Labs, and drive broad adoption of useful innovations across the UN System.

Global Pulse functions as a network of innovation labs where research on Big Data for Development is conceived and coordinated. We partner with experts from UN agencies, governments, academia, and the private sector to research, develop, and mainstream approaches for applying real-time digital data to 21st century development challenges. Our strategy includes:

1. **Research & Development:** Conducting research to discover new proxy indicators in digital data that can help us improve how development progress is tracked, and how we identify impediments to population wellbeing. This work will help us to develop a toolkit of the most effective methodologies and technology tools.
2. **Big Data Partnerships:** Forging partnerships with companies and organizations that have the data, technology and analytical expertise needed for the success of "Big Data for Development" research and advocacy. We work with companies to determine how they can effectively contribute data tools and expertise to the public good.
3. **Pulse Lab Network:** Working with UN Member States to establish an integrated, global network of Pulse Labs, where researchers work to prototype and pilot approaches at country level. Successful data analytics methodologies and technology tools that Pulse Labs develop would be widely shared for adoption by institutional partners.

Global Pulse is working to design an approach for harnessing big data and real-time analytics for monitoring development progress, emerging vulnerabilities and overall population well-being of the populations the United Nations serves.

WHO WE ARE

Our team of data scientists, open source hackers, and international development experts functions the way an R&D lab does: asking questions, formulating and testing hypotheses, building prototypes and collaborating with partners within and outside the United Nations to develop methods for harnessing real-time data to gain a real-time understanding of human well-being. We are supported by a network of volunteers from academia, the private sector, and the open-source tech community.

We currently have Pulse Labs in three global locations. Learn more about our teams in Jakarta, Kampala and New York.

Values & Principles

- **Promote Open Innovation** We develop our tools through an open and iterative process of design and development, driven by the needs of our users. We emphasize utility, usability, and accessibility for all.

- **Create Open-Source Technology** The technologies we create are free, open source, extensible, and based on open data standards.

- **Work Collaboratively** We build on existing initiatives, partner whenever possible on new projects, and innovate only where we must.

- **Democratization of Ideas** We believe the most valuable innovations will be created in countries where the needs are greatest.

- **Time-Sensitive Research Parameters** We address the challenges before us quickly, through experimentation, rapid prototyping, and field tests; and we view both successes and failures as opportunities for learning.

- **Shared Knowledge** If information is power, shared information is even more powerful.

- **Value Individual Privacy** We recognize that not everything can be shared. We give our users the tools they need to protect sensitive information.

MEET PULSE LAB JAKARTA

Our team works closely with UN agencies and government counterparts in Indonesia to get a better understanding of the policy and implementation issues they are grappling with and explore whether new data sources can provide useful insights that support decision making. We are not interested in research for research sake, but on results that benefit development. For this reason, our team comprises a mix of skills: data crunchers and geeks, but also monitoring and evaluation, partnership building, policy and innovation and change management experts. Our typical working day can include defining research questions with partners, identifying ways to embed real time monitoring in existing programs, scanning the horizon for new approaches and research, building partnerships with local universities, running trainings on big data, and much more!

Our core team is working closely with partners from academia, civil society, the private sector to make sure that we incorporate a variety of perspectives in our analysis and our projects benefit from multi-disciplinary expertise.

MEET PULSE LAB KAMPALA

Our team works closely with UN agencies and government counterparts in Uganda to get a better understanding of the policy and implementation issues they are facing with and explore how new data sources can provide useful insights that support decision making. Our focus is on research that yields new approaches and methods to benefit development and supplement the information and insights that policymakers draw upon in their work. For this reason, our team comprises a mix of skills including a data scientist, data engineer, partnerships manager and a research coordinator.

Our typical working day can include defining research questions with partners, analysing data, identifying ways to embed real time monitoring in existing programs, building partnerships with local universities and much more. Our core team is working closely with partners from academia, civil society, the private sector to make sure that we incorporate a variety of perspectives in our analysis and our projects benefit from our multi-disciplinary expertise.

MEET PULSE LAB NEW YORK

Our team in New York works closely with UN agencies and government counterparts to get a better understanding of the policy and implementation issues they are grappling with and explore how new data sources can provide useful insights that support decision making.

Our team comprises a mix of skills: data scientists, data engineers and data analysts but also a social scientist, legal and privacy expert, a partnerships manager and communications expert. This multidisciplinary approach helps us innovate and test new approaches that, if successful, can be scaled up and mainstreamed by UN partners. All our research projects, including findings and methodologies are shared via the research section of the website.

UN Global Pulse also has a Fellowship Program for current PhD students or post-doctoral researchers. Fellows come and work with us on their own research question, we host them here in New York and Pulse Lab Kampala and Jakarta also host their own fellows.

Our typical working day can include testing new tools for analysing data, defining research questions with UN partners, identifying ways to embed real time monitoring in existing programs, scanning the horizon for new approaches and research, building partnerships with local universities or speaking at conferences and seminars to share our approach.

STRATEGY AND ROADMAP

Global Pulse's objective is to foster development of the analytical, technological and organizational capacities that decision makers need to access and utilize new digital data sources and real-time analytics. This will facilitate faster, greater understanding of how vulnerable communities are being affected, even as a crisis or shock is still unfolding. It will lead to more agile, precise and efficient policy action.

Global Pulse's goal is widespread public sector uptake of the methodologies and technologies needed to incorporate real-time information into work norms. As such, the initiative is designed as a temporary scaffolding to promote awareness and the benefits of Big Data as a public good. We are working to build communities of practice, catalyze innovations in real-time monitoring, and accelerate adoption of proven solutions across development practice.

To achieve these objectives, Global Pulse's strategy includes:

1. **Research & Development:** Conducting research to discover new proxy indicators in digital data that can help us improve how development progress is tracked, and how we identify impediments to population wellbeing. This work will help us to develop a toolkit of the most effective methodologies and technology tools.
2. **Big Data Partnerships:** Forging partnerships with companies and organizations that have the data, technology and analytical expertise needed for the success of "Big Data for Development" research and advocacy. We work with companies to determine how they can effectively contribute data tools and expertise to the public good.
3. **Pulse Lab Network:** Working with UN Member States to establish an integrated, global network of Pulse Labs, where researchers work to prototype and pilot approaches at country level. Successful data analytics methodologies and technology tools that Pulse Labs develop would be widely shared for adoption by institutional partners.

PROJECT GOALS

Global Pulse's role as an innovation lab is to encourage and facilitate research and development (R&D) that leads to successful approaches for harnessing big data to understand changes in human well-being in real-time. The development of a new set of technology tools, partnerships and capacities - designed to complement existing data-gathering and analysis methods – will contribute to improved global development outcomes in three ways:

1. **Enhanced Early Warning:** Earlier detection of anomalies, trends and events allows earlier response to emerging crises.
2. **Real-Time Awareness:** A more accurate and up-to-date picture of what a population needs and wants can lead to better, more effective programme planning and implementation.
3. **Real-Time Feedback:** Understanding sooner where needs are changing or are not being met will allow for rapid, adaptive course correction in development programmes and current policies—in essence making impact evaluation and response more agile.

KEY MILESTONES

2010

- **Consultations and Program Design.** Series of workshops and consultations with UN Agencies, Member States, and thought-leaders in data science and development innovation communities to inform Global Pulse initiative's strategy and blueprint.
- **Voices of the Vulnerable Reports.** Publication of two high-level reports highlighting plight of most vulnerable communities in the wake of the global economic crisis.
- **Lab Location Scoping.** Lab Location Scoping. Scoping of first 2 potential pulse lab locations, establish preliminary presence in Kampala.
- **Strategic Plan for 2 Pulse Labs.** With the support of donors, UN agencies, and the Governments of Indonesia and Uganda, Global Pulse completes the scoping phases and develops strategic plans for the launch of the first two Pulse Labs in the network.

2011-2012

- **5 Proofs of Concept.** Research team completes series of 6 joint research projects with private sector and academia demonstrating that new sources of data may be combined with traditional data to speed up analysis of emerging vulnerabilities.
- **9 Interagency Research Projects.** Global Pulse hosts RIVAF Partners Conference ("Toward a Real Time Understanding of Emerging Vulnerability") that brought together UNDP, UNFPA, UNICEF, UNWTO, ILO, UNSPIDER, WFP, UNODC and UNESCO to share key findings of 9

projects Global Pulse funded investigating the impacts of the 2009 crisis on vulnerable populations.

- **HunchWorks Prototype.** With more than 300 pages of designs, and with tremendous external support from volunteer software developers, Global Pulse completes the first working prototype of Hunchworks, the first open source tool in the Global Pulse Technology Toolkit.
- **Strategic Plan for 2 Pulse Labs.** With the support of donors, UN agencies, and the Governments of Indonesia and Uganda, Global Pulse completes the scoping phases and develops strategic plans for the launch of the first two Pulse Labs in the network.
- **Launch of Pulse Lab Jakarta.** Pulse Lab Jakarta begins applied research by conducting a series of proof-of-concept projects in collaboration with UN and government partners, and sharing findings and any resulting technology prototypes.

2013-2014

- **Launch of Pulse Lab Kampala.** Pulse Lab Kampala begins applied research by conducting a series of proof-of-concept projects in collaboration with UN and government partners, and sharing findings and any resulting technology prototypes.
- **Applied Research & Development.** Global Pulse establishes a mechanism for joint research that brings UN agencies with programmatic needs together with cutting-edge data mining partners in the private sector and academia.
- **"Big Data for Development" Partnerships.** Global Pulse forges global and country-level partnerships with private sector partners with the requisite technology, expertise, and data to contribute to R&D and the Technology Toolkit.
- **Data Philanthropy Initiative.** Through consultations and partnerships, development of a global strategy for convening private sector partners to develop a model for safely sharing real time data from the private sector support evidence-based policy action.

2015-2016

- **Pulse Lab Network.** With at least 3 Pulse Labs launched, labs are sharing analytical methodologies and key innovations in relevant technologies to support institutional partners in the adoption of real-time data into their decision-making and monitoring.
- **Real-Time Monitoring Framework.** Building from continued joint research on real-time monitoring with governments, UN agencies, private sector and academia, publishes compilation of methods papers.
- **Pulse Lab Cookbook.** Cookbook, capturing lessons and best practices in analysis, technology innovation, community engagement and partnerships to support government use of Big Data for real-time development monitoring and planning.
- **Technology Toolkit.** Integrated suite of free and open source technology tools for data collection, analysis, and decision support made available to the global community.

- **Data Philanthropy Network.** Global Pulse assembles a global network of public and private sector partners sharing data through a secure network to support real time tracking of development.

RESEARCH

Inventory: Mobile Phone-Based Data Collection Projects (2010)

The use of mobile phones for quick-time data collection is proliferating around the world. To get a better understanding of the scale and scope of these new data collection efforts, the Global Pulse initiative partnered with MobileActive.org, a global network of people using mobile technology for social change, to conduct a survey of present and planned mobile data collection efforts. The survey results will help identify new, quick-time data sources that could be valuable to the Global Pulse initiative as sources of structured data (that can be used within a statistical framework to detect early impacts and changing vulnerabilities) or unstructured/semi-structured data (that could be fed into text mining engines for detection of events and patterns of interest). The first findings of the global survey have been compiled in an inventory. The inventory is a living document that will be regularly updated as the Global Pulse Team becomes aware of new projects. If you are managing a mobile data collection project and you would like to have it featured in the inventory, please contact us.

Question Box Analytics (2010)

Global Pulse has partnered with Question Box to test the potential of using search queries to identify emerging risks amongst vulnerable populations at a local level. The findings of this analytical assessment will be used as a test case for real time query based tracking of potential risks to vulnerable populations. In the long term, monitoring of anonymised query data may be able to provide local and macro indicators of emerging risks to the vulnerable. Question Box is a facility that allows people who generally have no access to the internet to make queries of a remote team. People can make the queries through a custom constructed device similar to a pay phone. Question Box Operators identify answers through internet search queries, databases of relevant local information and answers to previous questions. The answer is generally quickly identified and operators then translate it back into the local language. To improve accessibility the service could be extended in the future to support SMS queries in addition to voice. Voice service is currently operational in Uganda and India and has received thousands of queries. Many of the queries relate to agriculture and health although there are a wide range of questions about employment, business, politics, sport and other interests. Global Pulse is analysing an anonymised database of question box queries to explore the timing, frequency and type of queries. This assessment should identify trends in the queries that

provide strong indicators about events occurring at a local level. These indicators will then be retrospectively compared against recorded events. It is expected that queries will be correlated with events and trends in the study communities. For example, an increase in queries about jobs may indicate rising unemployment, queries about particular crop symptoms may be correlated with an outbreak of agricultural disease and/or queries about government assistance may indicate increase household vulnerability.

This non-traditional form of data capture also has additional advantages including building local resilience by assisting vulnerable people with targeted information and avoiding incentives to "game" the system that may exist in compensated surveys. Finally in the long term communities may even be willing to pay for the service, making it self-sustaining.

Rapid Impact and Vulnerability Analysis Fund (RIVAF) 2010-2011

Under the umbrella of the Rapid Impact and Vulnerability Analysis Fund (RIVAF), Global Pulse supported a series of studies and research projects, which were undertaken by eleven UN agencies, between 2010 and 2011. The topics represented in these eight research projects, present a unique view of how the global economic crisis has affected a variety of different sectors across the globe.

The work detailed in the short executive summaries in this final report truly showcases the work of the United Nations—not only in its geographic spread, but in its cross-sectoral scope. Taken as a whole, the results provide a sobering lesson in the variety of ways that vulnerability manifests in the wake of economic shocks—from increasing crime to the introduction of informal school fees to an increased willingness to undertake unsafe migration. The findings also highlight the variety of populations who are affected—from those dependent on remittances in Colombia, to migrant workers in the tourism sector in the Maldives, to cattle farmers in Ethiopia. The impacts on any one of these disparate populations can provide only a snapshot in time; but taken together, they start to form a fuller picture, one that illustrates the effects of the crisis on "main streets" globally.

The Rapid Impact and Vulnerability Analysis Fund (RIVAF) was launched by the Global Pulse Initiative in December 2009 with funding from the UK's Department for International Development (DfID) and the Government of Sweden. The fund supports innovative, real-time data collection and analysis to help develop a better understanding of how vulnerable populations cope with impacts of global crises.

The RIVAF report includes summaries from eight research projects (individual papers can be accessed via the links below):

- UNDP/UNICEF: "Monitoring Household Coping Strategies During Complex Crisis"
- UNESCO: "Monitoring the Implications of the Global Financial Crisis on Primary Schools, Teachers and Parents in 12 Countries"
- UNFPA: "Impacts of the Financial Crisis on Health and Poverty in 3 Countries (Colombia, Jordan and Ethiopia)"
- UNICEF/WFP: "Food and Nutrition Security Monitoring and Analysis Systems: A Review of Five Countries (Indonesia, Madagascar, Malawi, Nepal and Zambia)"
- UNODC: "Monitoring the Impact of the Economic Crises on Crime"
- UNODC: "Smugglers and Vulnerable Migrants in Central America and Mexico"
- UNOOSA: "A Visual Analytics Approach to Understanding Poverty Assessment through Disaster Impacts in Latin America and Africa"
- UNWTO/ILO: "Economic Crisis, International Tourism Decline and its Impact on the Poor"

Findings from the individual research projects revealed fascinating insights, painting a diverse picture of the impacts of global crises. Some highlights include:

- Borrowing money is one of the two most prevalent coping strategies, with 37.34% of households reporting borrowing to meet various expenses. (UNDP/UNICEF research)
- Primary schools appear to be prone to charging user fees during times of economic difficulty despite global advocacy and national legislation on free primary education. (UNESCO research)
- While the global economic crisis seemed to have had no significant impact on the overall number of illegal migrants. What did change was the level of vulnerability of illegal migrants—not only did the cost of migration increase, but the passage itself became more risky. (UNODC research)
- While the tourism sector proved in some ways to be more resilient than other sectors of the economy, case studies revealed distinct vulnerable groups within the tourism sector— for example, in the Maldives, this included foreign workers; in Costa Rica it was primarily women and low-skilled workers; and in Tanzania it was workers with dependents who were hardest hit. (UNWTO & ILO)

Unemployment Through the Lens of Social Media (2011)

Description:
This project investigates how social media and online user-generated content can be used to enrich the understanding of the changing job conditions in the US and Ireland by analyzing the moods and topics present in unemployment-related conversations from the open social web and relating them to official unemployment statistics. Two specific questions were addressed:

can online conversations provide an early indicator of impending job losses, and can they help policy makers enrich their understanding of the type and sequence of coping strategies employed by individuals?

Partner:
SAS (www.sas.com)
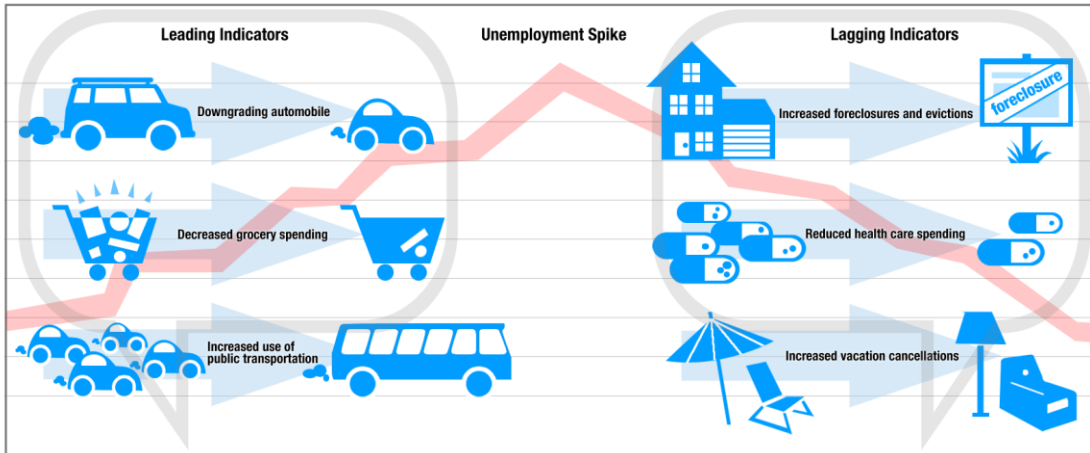
Analysis of social media using SAS shows increases in chatter about certain topics that are leading and lagging indicators of a spike in unemployment.

Project Overview:
This research investigates whether and how social media and other online user-generated content could enrich understanding of the effect of changing employment conditions. The primary goal is to compare the qualitative information offered by social media with unemployment figures.

To this end, we first selected online job-related conversations from blogs, forums and news from the United States and Ireland. For all documents, a quantitative mood score based on the tone of the conversations—for example happiness, depression or anxiety— was assigned. The number of unemployment-related documents that also dealt with other topics such as housing and transportation was also quantified, in order to gain insight into populations' coping mechanisms.

Analyzing half a million blogs, forums and news sites, two years of social media data from the US and Ireland were analyzed for references to unemployment and how people were coping.

The data was analyzed in two primary ways. First, the quantified mood scoring was correlated to the unemployment rate to discover leading indicators that forecast rises and falls in the unemployment rate. For example, the volume of conversations in Ireland categorized as showing a confused mood correlated with the unemployment rate with a lead-time of three (3) months.

Second, the volume of documents related to coping mechanisms also showed a significant relationship with the unemployment rate, which may give insight into the reactions that can be expected from a population dealing with unemployment. For example, the conversations in the US around the loss of housing increased two (2) months after unemployment spikes.

Overall, in this initial research, SAS and Global Pulse have underlined the potential of online conversations to complement official statistics, by providing a qualitative picture demonstrating how people are feeling and coping with respect to their employment status.

The Dashboard allows the user to investigate the volume of conversations around unemployment as well as the coping mechanisms that are being discussed in relation to unemployment. The Dashboard also identifies time relationships between unemployment, conversation moods, coping mechanisms and various macroeconomic indicators, which allow the user to see patterns and predictions.

Monitoring Food Security Issues Through News Media (2011)

Description:
This project finds emerging trends and organizes thematic clusters in news related to food security issues in French language media from the last 8 years using automated text analysis, semantic clustering and networks theory.

Partners:
The Complex Systems Institute of Paris Île-de-France (ISC-PIF, www.iscpif.fr) - CAMS, CNRS, CorText, CREA, Ecole Polytechnique, ESIEE, Formism, IFRIS, INRA-SenS - and The Institute For Research, Innovation and Society (IFRIS, www.ifris.org)
Team:
David Chavalarias, Jean-Phillipe Cointet, Lise Cornilleau, Tam Kien Duong, Andreļ Mogoutov, Camille Roth, Thierry Savy, Lionel Villard

Project Overview:
This project explored whether is possible to track and comprehend thematic shifts in media attention through the automatic analysis of news articles.

To do so, we analyzed how the Francophone media reported on food security issues over the past 8 years. More than 20,000 related articles published between 2004 and 2011 were selected using an ontology of food security related terms. The contents of these articles were analyzed and organized into thematic clusters, which can be traced back to specific news articles or aggregated in a big picture representation. Semantic clustering methodologies and networks theory were used to automatically produce those clusters. The news items were also geo-located, enabling us to map how a given theme or issue is distributed over the world. Moreover, themes identified at successive time steps were reconnected into streams of content. A stream visualization illustrates how topics are articulated through time.

An interactive online interface visualizes these maps, themes and news entries to answer questions such as: Is an issue - concerning, for example, the impact of climatic change on food security - attracting more attention with time? How does a specific issue relate with contiguous subjects (use of biofuel, for example)?

The interactive platform can be explored online at: http://pulseweb.veilledynamique.com

Tubes with the same color form hyper-streams of a common topic. The analysis shows that as the 2008 global economic crisis unfolded, news coverage shifted from a focus on humanitarian issues (red) to food price volatility (blue), whereas in 2011, the news focus has shifted to social unrest (magenta).

Global Snapshot of Wellbeing - Mobile Survey (2011)

Description:
The objective of this ongoing collaborative project is to experiment with new tools that can replicate the standards of traditional household surveys, in real-time on a global scale. In particular, Jana and Global Pulse are testing the potential of deploying mobile surveys, over SMS, in multiple countries. The project will also explore the feasibility of conducting longitudinal surveys over time and the possibility of reaching underrepresented populations that have access to mobile phone technology.

Partner:
JANA (www.jana.com)

Project Overview:
Household surveys have long been the gold standard for collecting data on populations, both for national statistics, impact assessment, and programme monitoring & evaluation. However, household surveys are burdensome and costly, and thus difficult to scale. Mobile-based surveys have already shown, in recent years, their potential to complement traditional data collection at the country-level.

For this project, Global Pulse partnered with Jana to conduct a survey at global scale.

In its network, Jana has over 2 billion mobile subscribers who have opted into a process where they will answer survey questions in exchange for a small amount of airtime. Since the majority of plans in the developing world and emerging markets rely on pay-as-you-go fees, this method has the potential to cover a great deal of the world's mobile users.

Global Pulse collaborated with Jana to develop a set of questions about well-being and interconnectedness that were translated into 15 languages, and sent over mobile phones directly to citizens more than 30 countries, collecting more than 90,000 answers between August and November 2011.The short video below (produced by Paul Butler, as a part of Data Without Borders) demonstrates the rapidity of data collection by showing the number of responses collected from July 25 to October 16, 2011.

Most of the questions were multiple-choice, with participants pressing a number on their keypad corresponding to an answer. Some of the questions were free-text, but expected a numerical answer (i.e. "How many days did you work in the last 7 days?"). Finally, a very few of the questions were free text and required a text answer (i.e. "If you had 15 USD, how would spend it?").

Below is a sample of the survey questions we asked, clustered by theme:

Questions about Well-Being
- How did you feel over the past 7 days?   *Answer Options: 1 - Great, 2 - Good, 3 - Bad, 4 - Neither good nor bad, 5 - Very Bad*
- How many days did you work in the past 7 days? *(Free Text)*
- Were you sick in the past 7 days? *Answer Options: 1 - No, 2 - Yes but did not need treatment, 3 - Yes and received treatment, 4 - Yes and needed medical treatment but did not receive it*
- If you had 15 USD, what would you spend it on? *(Free Text)*

Questions about Interconnectedness:
- How much money did you spend on airtime in the past 7 days? *(Free Text)*
- In the past 7 days, did you communicate with friends and family using: *Answer Options (please indicate all that apply): 1 - Face to Face, 2 - Mobile (voice), 3 - Mobile (SMS), 4 - Landline, 5- Internet (email), 6 - Internet (social networks)*
- In the past 7 days, did you communicate with anyone (please indicate all that apply:*Answer Options: 1 - Within walking distance, 3 - In your city/village, 4 - In your province/state, 5 - In your country, 6 - Out of your country*
- Have you ever purchased goods or transferred money using your mobile phone? *1- Yes, 2- No*
- Have you ever purchased goods or services over the Internet? *1- Yes, 2- No*

--

A visualization of a partial subset of the answers was included in the "PopTech: World Rebalancing" iPad app earlier this year.

The PopTech app provided an opportunity for Global Pulse to begin visualizing some of the results of the survey. While the answers that were visualized in the app are not a representative sample (in fact, some countries only yielded less than 50 responses), this

exercise was an excellent opportunity to start thinking creatively about how to present the results of the larger survey, in an interactive format.
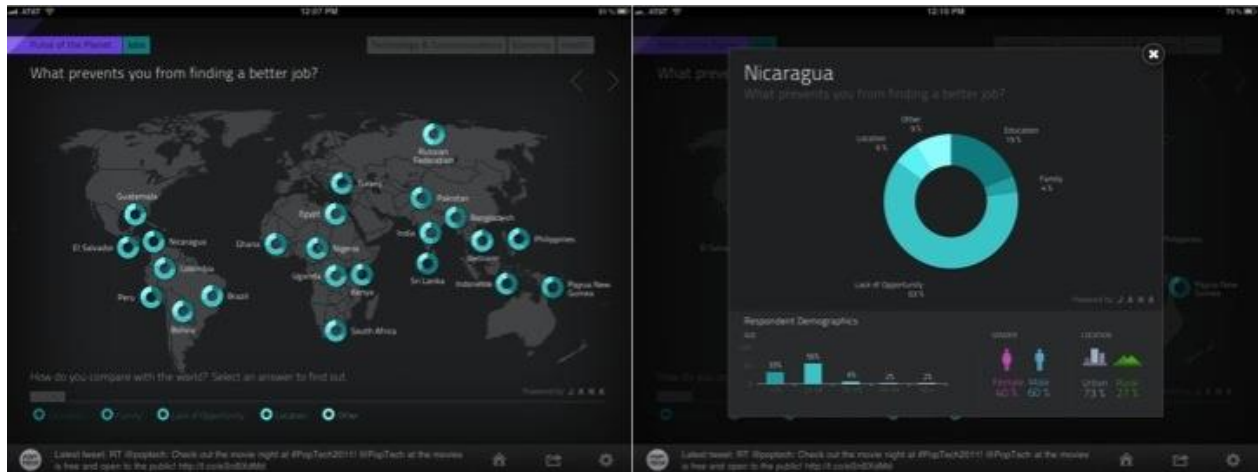
A next step for the Global Pulse/Jana collaboration, is to perform a more in-depth analysis on the results to date, and eventually to conduct similar surveys over time in order to map the evolution of trends in well-being over time, continue to establish demographic baseline data (including information about economic status, gender, age, literacy, etc.) and to see what works, what doesn't work and why.

Daily Tracking of Commodity Prices: the e-Bread Index (2011)

Description:
This project investigates and shows how scraping online prices could provide real-time insights on price dynamics, focusing on the case of bread in 6 Latin American countries.

Partners:
PriceStats and the Billion Prices Project at MIT
Project Overview:
The objective of this project was to investigate and show how scraping the web for online prices could provide real-time insights on price dynamics, focusing on the case of bread. The data is collected in Argentina, Brazil, Chile, Colombia, Uruguay and Venezuela using a scraping software that records, on a daily basis, the price of bread sold or advertised in online supermarkets. Then the daily inflation rate of bread for each country is calculated. Then, subsequently, the average daily inflation rate for each of the bread products is estimated, and an index aggregating those daily changes, is built by country. All the bread products are equally counted and the resulting inflation is the geometric average of their daily price inflation.

| ARGENTINA | BRAZIL | CHILE |
| COLOMBIA | URUGUAY | VENEZUELA |
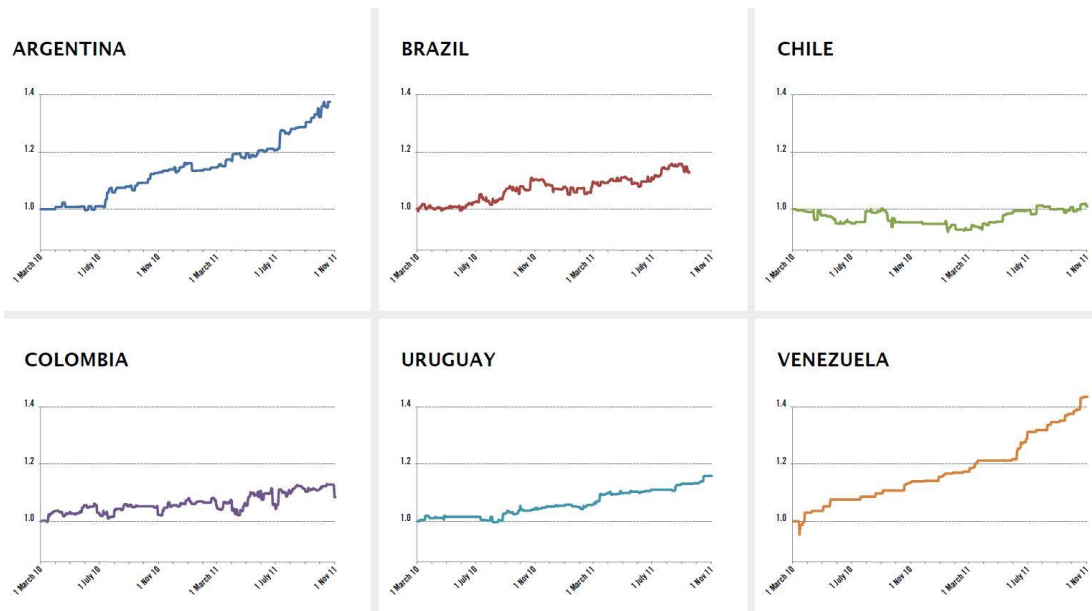
The daily e-bread index - built on data extracted from online retailers - from 1st March 2010 to 1st November 2011 shows the change in prices of bread (normalized to the 1st March).

Argentina and Venezuela indicators show a quasi-monotonous increase in prices, while Brazil, Chile, Colombia and Uruguay tended to fluctuate, somewhat reflecting the behavior of wheat prices. Overall, the e-bread indexes do indeed follow a similar trend as the official consumer price index for the food basket; for example, Uruguay has a very high correlation with the specific CPI for bread and cereals. The difference is that online prices can be obtained daily whilst consumer price indices in most countries are only published on a monthly basis. Overall, this case study shows the potential offered by online marketplaces to scrape prices in real time and construct highly reactive inflation measures that complement other existing sources of data.
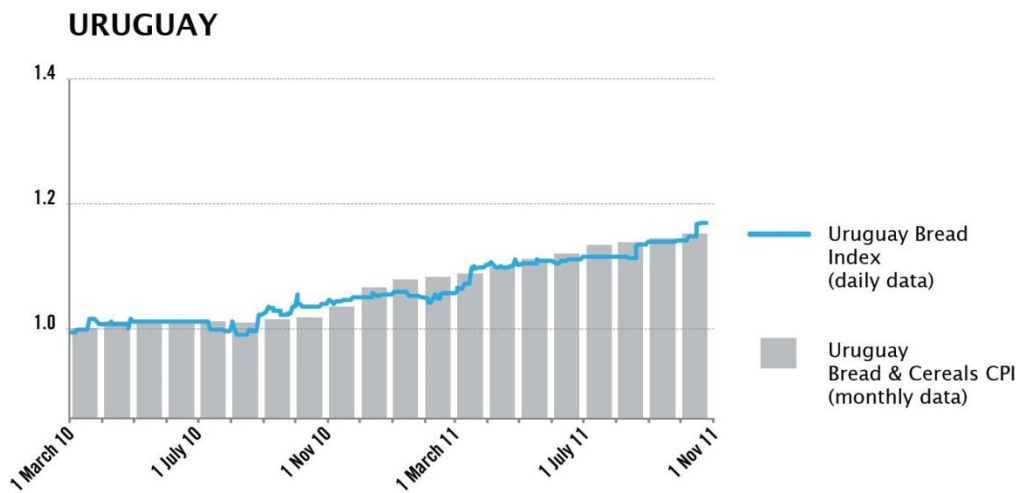
The e-Bread index for Uruguay strongly correlates with the Bread & Cereals consumer price index.

Twitter and Perceptions of Crisis-Related Stress

Description:

This research identifies and quantifies discussion themes in Twitter data in order to investigate what indicators can help understand people's perceptions and concerns around food, fuel, finance and housing in the US and Indonesia. *Partner***:** Crimson Hexagon (www.crimsonhexagon.com)

Overview:

The purpose of this research project is to determine which indicators might be present in social media data that could shed light on how populations cope with global crises, such as commodity price volatility or the continuing global economic crisis. In this investigation, the analysis was limited to publicly available data from Twitter for July 2010 through October 2011 in Javanese/Bahasa Indonesia and English. The topics of focus included the affordability/availability of food, fuel, housing and loans. By classifying a populations' tweets into several categories associated with relevant topics, it was possible to perform quantitative analysis to better understand populations' concerns:

- detecting anomalies such as spikes or drops in the number of tweets about particular topics (e.g. comments about power outages in Indonesia or student loans in U.S.),
- observing weekly and monthly trends in Twitter conversations (e.g. discussions around debt in U.S.),
- finding patterns in the volume of particular topics over time (e.g. discussions around housing in U.S.),
- comparing the proportions of different sub-topics to understand shifts in trends over time (e.g. the ratio of tweets about formal loans vs. informal loans in Indonesia),
- or relating trends in Twitter conversations with external indicators (e.g. conversations around the price of rice in Indonesia mimicking the official inflation statistics).
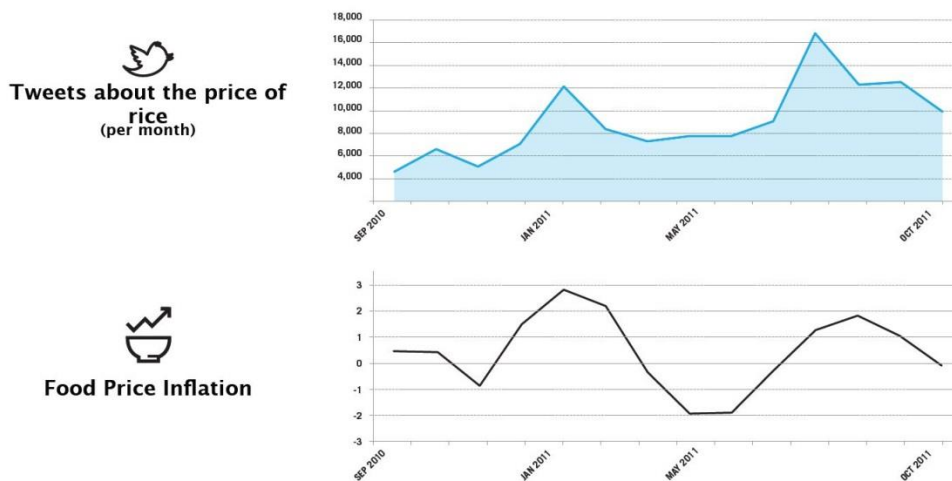
The number of tweets discussing the price of rice in Indonesia over the last year follows a similar function as the official inflation statistics for the food basket.

This research has confirmed that Twitter data can be useful for understanding the immediate worries, fears and concerns of populations, but at the same time, the research suggested that it is a poor source of data for gauging people's *long term* aspirations. There are several remaining challenges, in particular that Twitter has a specific culture and demographic which needs to be better understood. Overall, this exploratory research shows some of the potential of Twitter data for exploring people's perceptions of crisis-related stress and suggests research lines and methodologies for further investigations.
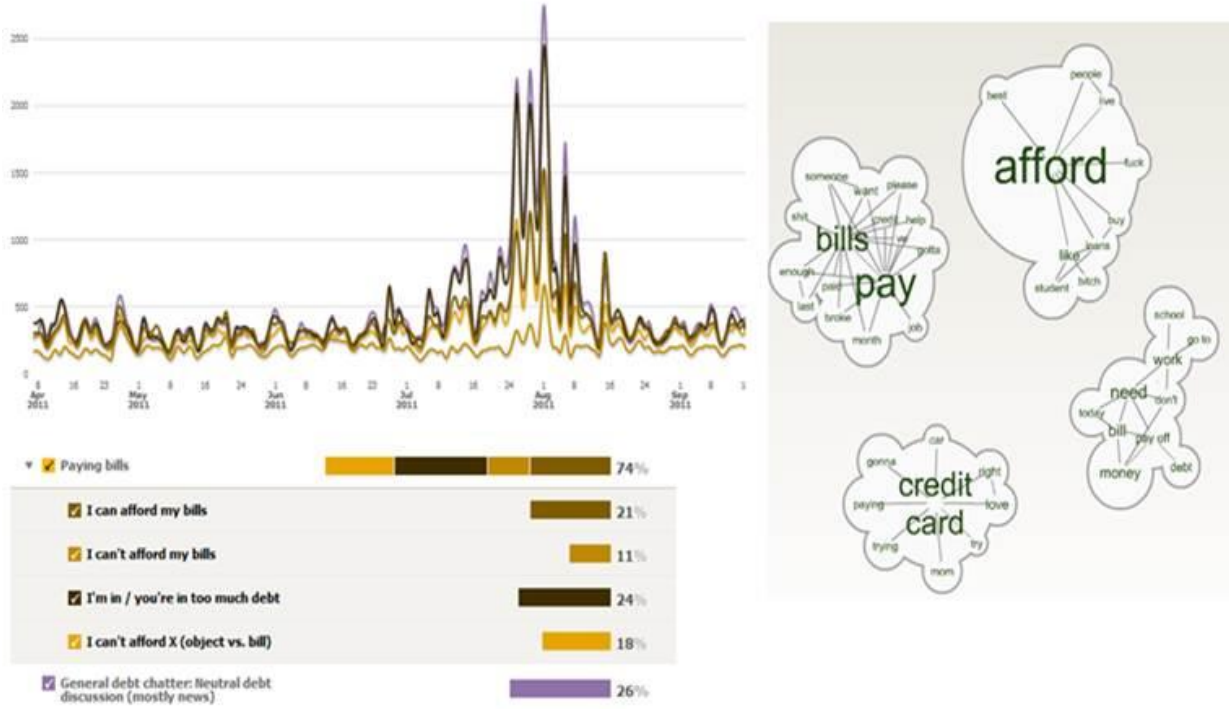
Conversations around finance in the US, modulated by a baseline weekly pattern of fewer discussions on the weekends, show an increase of conversations from the 15th July to the 15th August motivated by the US debt ceiling debate.

"Big Data for Development: Opportunities & Challenges" - White Paper (2012)

Global Pulse White Paper "**Big Data for Development: Opportunities & Challenges**" (published May 2012) aims to highlight the opportunity, as well as some of the main concerns and challenges, raised by utilizing new, digital data sources in the field of international development, as concretely and openly as possible, and to suggest some ways forward. The report is structured to foster dialogue around some of the following issues:

- What types of new, digital data sources are potentially useful to the field of international development?
- What kind of analytical tools, methodologies for analyzing Big Data have already been tried and tested by academia and the private sector, which could have utility for the public sector?
- What challenges are posed by the potential of using digital data sources (Big Data) in development work?
- What are some specific applications of Big Data in the field of global development?
- How can we chart a way forward?

Understanding the Post-2015 Global Conversation through Big Data

Project Overview:

At the turn of the millennium the United Nations launched a global development framework known as the Millennium Development Goals (MDGs). Eight in total, with accompanying indicators and targets, the MDGs provided a roadmap to halve poverty worldwide by 2015.

As we near the 2015 deadline, the UN, in partnership with civil society and the private sector is undergoing a process to reflect on the MDGs as well as develop a new global framework Post-2015.

As part of the Post-2015 process, the UN convened a panel of world leaders and experts known as the High Level Panel, facilitating 9 global thematic consultations and over 50 country consultations, running a global survey to ask people what they think about key issues, and leading a Global Conversation which will be an effort to allow citizens and other stakeholders to participate and contribute ideas. The High Level Panel Report called for a 'data revolution' to improve the real-time monitoring and progress of the Post-2015 development goals.

As a supplemental exercise, Global Pulse is working with the UN Millennium Campaign to demonstrate the potential of big data and visual analytics as a truly 21st century way of

unearthing which global development topics people talk the most about online, even if they are unaware of the official Post-2015 process taking place.

We mine and visualize public online chatter (tweets, and other relevant social media content) to show what key issues relevant to the Post-2015 framework people talk about (and at what volume).

For each of these topics, a relevant set of keywords in English was established and used to analyze the Twitter firehose. Thus far, in the research, which is still ongoing, over 87 million tweets have been analysed, in four languages, spanning 193 countries.

The volume of conversations around each of these topics was then compared, and preliminary results, shared on the Post-2015 initiative's website, suggest that "Better job opportunities" and "A good education" are the most consistently talked about, while "Freedom from discrimination and persecution" spikes in the overall numbers when particular events happen.

Project Partner:
The UN Millennium Campaign

Landscaping Study: Digital Signals & Access to Finance in Kenya

Description
In emerging markets, eight out of ten small businesses cannot access the loans they need to grow. USAID's Development Credit Authority (DCA) helps small businesses to access capital. The goal of this collaboration between UN Global Pulse and USAID is to explore how big data could support the work of USAID's Development Credit Authority. Kenya has become an established tech leader in Africa in recent years – generating greater volumes of digital data as a result. The goal of this study is to explore what new sources of digital data, and methods for analysis, could be helpful in answering the question: "What barriers to accessing loans do small businesses in Kenya face?". Accordingly, the final report (below) paints a picture of the big data landscape in Kenya, shows preliminary findings, and lays the groundwork for further investigation.

Project Partner: USAID Development Credit Authority

Project Overview

The landscaping study was conducted in three steps:

1.   Background research on the digital media landscape in Kenya

2. Developing a taxonomy (key words) related to loans, as commonly used in Kenya, through surveys and field research
3. Using a variety of tools to analyze trends in social media and online search

Project Summary

The study began by looking at baseline data for ICT and social media usage rates in Kenya. By interviewing a sample of DCA's clients – farmers from the region around Nakuru and Nairobi – Global Pulse created a taxonomy of colloquial language used in relation to accessing loans and finance. Global Pulse's analysts then used this taxonomy to mine available data from Twitter and Google search trends, to understand online conversations and information-seeking behavior.

It's clear that in Kenya's current digital landscape, the quantity of relevant social media data restricts its utility. For example, the study found that in July 2013 there was an average of only about 10 tweets related to "loans" on Twitter, per day. The relatively small numbers of relevant digital signals available indicate that in the short-term, social data can likely only provide supplementary insight about barriers to finance in Kenya. For instance, online analysis could be useful for revealing early themes or trends, to inform the topics of focus groups to validate. Despite the relatively small numbers of contextually relevant tweets available in 2013, there is an emergence of a Kenya-specific Twitter culture. In particular, Twitter is being used to seek, access and share information about loans, especially mobile loans. The introduction of a savings and small loans service called M-Shwari to M-Pesa customers in Kenya has already sent ripples through Kenyan social media.

The study also conducted some initial tests with Google to understand finance related information-seeking behavior in Kenya. Events that are broadly shared (i.e. the beginning of a school semester or elections) are clearly visible in the data. This could mean that other systemic, or anomalous, events will be visible as well. However, the analytical tools made available by Google for Kenya data are limited - currently it is only possible to know the relative volume of searches, rather than access real numbers.

In addition to this analysis, the study looked at the digital footprint of the Development Credit Authority, its clients and priority sectors, and offers a menu of what other relevant sources of digital data might exist in Kenya, and how the data could potentially be accessed.

Project status: Landscaping study completed

Analyzing Online Content for Insight on Women and Employment in Indonesia

Project Objectives

The project aims to analyze online content (such as news articles, blogs, and social media) to gain new insight on public perceptions about women's role in the workplace, their conditions of employment and obstacles to equality and equal opportunities in the workforce. This would support the government and ILO Indonesia in developing strategies to improve women's access to employment and monitor the impact of major advocacy efforts on the topic.

Project Partners

Pulse Lab Jakarta, International Labour Organization (ILO-Indonesia), Ministry of National Development Planning (Bappenas)

Project Context

Women's economic empowerment is integral to human development in Indonesia. However, women continue to face discrimination in the hiring process as well as in the workplace across several Indonesian industries. This perception is reinforced with discriminatory workplace practices; women earn approximately 25 per cent less than men, workplace structures are inflexible and make work-life balance unmanageable for many female workers, and violence in the workplace is still common, especially among migrant workers. In the informal economy, women often work unpaid where working conditions and job security are even poorer.

Monitoring these issues can be challenging. Traditional monitoring is done via surveys, which are typically costly, and only available at limited points in time. Big Data sources and new technologies have the potential of providing new insights on populations' behaviors in real-time, and can supplement traditional monitoring methods by informing the design and targeting of rigorous surveys.

In that context the ILO Jakarta Office has requested that Pulse Lab Jakarta support them in analyzing online perceptions of barriers to women's access to employment and obstacles to decent work as a complement to traditional monitoring methods.

Project Status

Active

Advocacy Monitoring through Social Data: Womens and Childrens Health

Project Objectives

First of the short-term research projects resulting from Global Pulse's recent call for research proposals, this analysis aimed at evaluating the impact of the *Every Woman Every*

Child movement since is launch in 2010, as well as determining whether there has been any change in general public awareness of issues related to children's and women's (particular mothers) health reflected in Twitter conversations.

Project Partners

Every Woman Every Child, United Nations Foundation

Project Overview

Using Crimson Hexagon's analytical tool Forsight, we were able to access and analyze an archive of all public tweets from September 2009 to July 2013. A taxonomy of relevant keywords was developed to identify messages related to women's and children's health (for instance, searching keywords such as "maternal health," "breastfeeding," "vaccination of children"). After searching the Twitter firehose for occurrence of the keywords, we trained a monitor to recognize relevant tweets. The resulting 14 million tweets about women's and children's health were then analyzed to identify spikes, trends and possible connections with real life events and campaigns.

Preliminary Findings

Close to 14 million tweets have been generated by the global public in the last 3 years related to the topics of children's and women's health
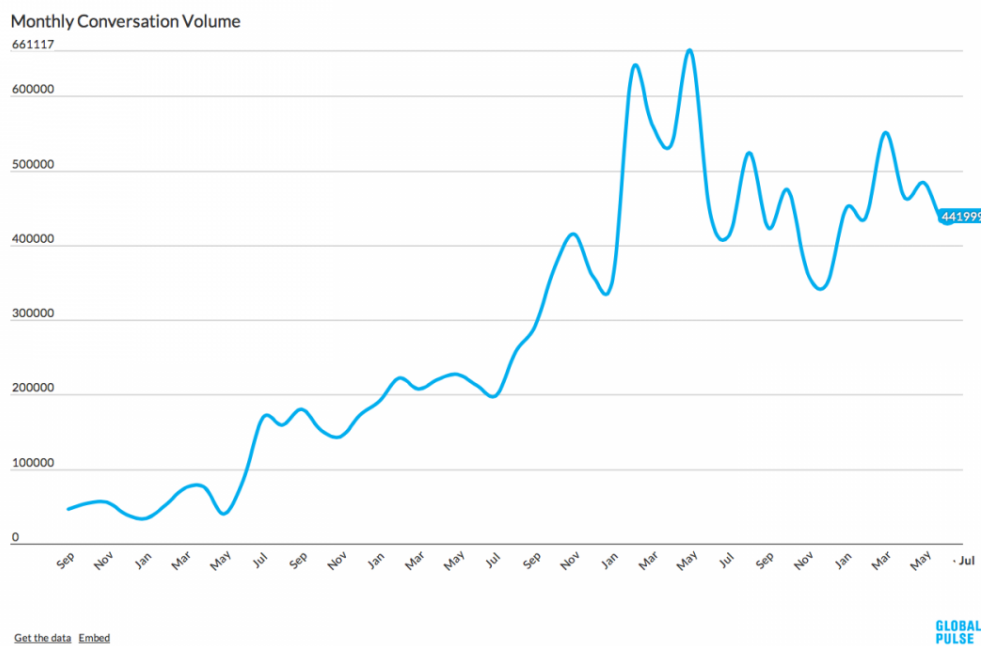
The trend analysis showed that, in the period of investigation, there has been an increase in English tweets related to women's and children's health of 1000%. (An important caveat to note, however, is that this is largely attributable to greater numbers of people using the Twitter service over the past two years. The next phase of our analysis will adjust for "Twitter inflation.") While content relevant to children's health represent the majority of tweets

throughout the period of investigation, the relative growth of the conversation about women's health over the past 18 months has been exponential.

The study also shed some insight on who is engaged and paying attention to the issues, from a demographic perspective. While the majority of users talking about the women's and children's health are women, the percentage of men tweeting about these topics grew through the period analyzed, at times surpassing the 50%.

The volume of tweets reflected real life events and initiatives on topic days. Specifically, spikes in conversations were observed in occurrence of World Aids Day, Mother's Day, International Women's Day and even at key moments such as the MDG summit and Women Deliver conferences, showing that advocacy efforts are working and the public mind-share and attention is being maintained on the issues.

Mining Indonesian Tweets to Understand Food Price Crises

Description
The objective of this research project was to monitor Twitter conversations in Indonesia to understand how conversation volume trends relate to world events. In particular, this project monitored food-price related tweets between January 2011 and December 2012 to see if variations in their volumes could be connected with food and fuel price inflation.

Project Partners
[Pulse Lab Jakarta](), [Indonesian Ministry for Development Planning]() & [Crimson Hexagon]()
Project Overview
The first step in the project was to create a taxonomy of relevant keywords and phrases in Bhasa Indonesia to extract relevant tweets, which included phrases related to the prices of food and fuel. As a second step, the researchers define categories in which to classify the relevant tweets, depending on the sentiment they express (i. e. "positive, " "negative, " "confused. "). A representative, hand-labeled sample of tweets was then used to train a monitor to classify the tweets in the correct category and detect the sentiment of new tweets being published in real time.

Finally, all relevant data was analyzed to provide a proportion of tweets related to each theme to determine the statistical pattern of conversation for each category. The general volume of relevant tweets, independent from the conversation, was analyzed as well, and three spikes in volume of tweets were observed in 2012, corresponding to three real-world events:

- **July 2012:** a global soybean price rise, which affected the prices of tempeh and tofu, two dishes made of soybeans consumed by many Indonesians as affordable protein-rich options.

- **March 2012:** an proposal by the Indonesian government that its fuel subsidy would be cut by 33%, which caused violent protests and raised concerns that food prices may eventually be affected.
- **November 2012:** approval of a law establishing a new food agency with policymaking authority with the aim of helping Indonesia reach self-sufficiency in staple foods, including rice and soybeans.

The initial research results showed that around the same time when these real-world events occurred, conversations related to food prices also spiked dramatically among Indonesian Twitter users, illustrating the potential value of employing regular social media analysis for early warning and impact monitoring.

Characterizing human behavior during floods through the lens of mobile phone activity

Global Pulse is collaborating with researchers from Spanish telecommunications provider Telefonica (Telefonica Research), the Technical University of Madrid, the UN World Food Programme (WFP) and the Government of Mexico on a joint research project exploring how anonymised call detail records (CDRs) could be used to understand populations behaviours and migration after floods.

Project Partners:

[Pulse Lab New York](#), [Telefonica Research](#), [Technical University of Madrid (UPM)](#), [World Food Program - Vulnerability Analysis and Mapping](#), [Digital Strategy Coordination Office of the President of Mexico](#)

Project Overview:

Natural disasters affect hundreds of millions of people worldwide every year. Emergency response depends on the availability of timely information such as the description of movements of affected populations. The analysis of Call Detail Records (CDR) derived from mobile phone activity provides new possibilities to characterize human behaviours during critical events. In this project, we will investigate the viability of using CDR data to understand the floodings occurred in the region of Tabasco (Mexico) in 2009, combining CDR data with other sources of information.

The ongoing retrospective analysis focuses on the floods that occurred in the Tabasco region in Mexico in 2009. The preliminary analyses of communication and migration patterns revealed particular data signatures occurring in the cell towers located in the most affected regions.

An impact map has been reconstructed using Landsat-7 images to identify the floods. Within this frame, the underlying communication activity and mobility signals in the CDR data have been analyzed and compared to rainfall levels extracted from data of the NASA-TRMM project. Therefore, we propose a new metric at the cell tower level derived from the variations of the number of users located in the tower, that allows to capture the abnormal activity patterns found in the most affected locations during the floods. The representativeness of the analysis has been assessed using census data and civil protection records. These early results suggest a high potential to trigger early warning and emergency management mechanisms, however a more extensive validation is required.

Status:

Preliminary exploration complete. Summary and findings report in progress.

Online Signals for Risk Factors of Non-Communicable Diseases (NCDs)

Global Pulse is collaborating with the World Health Organization (WHO) exploring how risk factors (e.g., tobacco, alcohol, diet and physical activity) of non-communicable diseases (e.g., cancer, diabetes, depression) could be inferred from big data sources as social media and online internet searches.

Project Partners

WHO Regional Office for Europe, Pulse Lab New York

Project Overview

70% of the reduction in coronary mortality can be attributed to population level reductions in the risk factors. Traditionally, monitoring of NCD risk has been conducted by cross-sectional surveys of the population which are costly and time-consuming, providing only a snapshot at a point in time. Big data sources as social media could provide faster and cheaper information on risks on an ongoing basis.

In an exploration project together with Microsoft Research, indices for each risk factor were built and tracked over time on Twitter data and internet search traffic for keywords such as 'stop smoking' were also analyzed. This project and a broader perspective of the potential of Big Data for NCDs was explored at a WHO technical conference held in Tallinn, Estonia in September 2013, and a research report for public dissemination is in development.

Status:

Preliminary exploration complete. White Paper and research report in progress.

Understanding parental perceptions of immunisation in real-time

Pulse Lab Jakarta is collaborating with WHO, UNICEF, Bappenas and the Ministry of Health to analyze online perceptions of immunization on social media platforms. Preliminary analysis shows that a number of sub-topics are discussed most on social media in Indonesia.

Project Partners:

Pulse Lab Jakarta, UNICEF Indonesia, WHO Indonesia, Indonesian Ministry of Health

Project Overview:

In the initial phase of this project, a taxonomy of nearly 1,000 keywords and phrases related to immunization and vaccination in Bahasa Indonesia were used as a filter to determine what volume of tweets were published in Indonesia on this topic over the last 24 months. Nearly 85,000 relevant tweets were filtered from a two-year archive of twitter data, commencing the beginning of 2012. For the analysis, fundamental social media analytics methods, such as volume dynamics, text analysis and influential user identification, were applied.

Preliminary analysis shows that the four primary topics of discussion related to immunization in Indonesian social media include: (a) new vaccines, (b) debates about the ethical and religious dimensions of immunisation, (c) discussion when there is news of outbreaks and (d) discussion of side effects of immunization. Analysis also identified the most influential sources for each of these topics, and the most popular words or phrases within each topic has also been initiated. This project should provide the baseline for an evaluation of the impact of advocacy campaigns, as well as developing guidelines for how social media monitoring can be used to inform the design and responsiveness of future advocacy campaigns.

Status:

Preliminary exploration complete. Summary and findings report in progress.