Big Data for Social Good

Nuria Oliver, PhD
Scientific Director
User, Data and Media Intelligence
Telefonica Research
6.8 billion subscriptions
96% of world’s population (ITU)

Mobile penetration of 120% to 89% of population (ITU)

More time spent on our phones than watching TV or with our partner (US and UK)

Emerging and developed regions
Digital footprints enable large-scale analysis of human behavior

May 19, 2011, 7:06 pm The Sensors Are Coming!
By NICK BILTON

Cell Phones as Sensors of Human Activity

Scientists want to put sensors into everyone's hands
Big Data from Cheap Phones

Collecting and analyzing information from simple cell phones can provide surprising insights into how people move about and behave—and even help predict future behavior.
HIGH LEVEL PANEL RELEASES RECOMMENDATIONS FOR WORLD’S NEXT DEVELOPMENT AGENDA

Eminent Persons from Around the World Call for a New Global Partnership to Eradicate Poverty and Transform Economies through Sustainable Development

The High Level Panel on the Post-2015 Development Agenda today released “A New Global Partnership: Eradicate Poverty and Transform Economies through Sustainable Development,” a report which sets out a universal agenda to eradicate extreme poverty from the face of the earth by 2030, and deliver on the promise of sustainable development. The report calls upon the world to rally around a new Global Partnership that offers hope and a role to every person in the world.

Wanted: A data revolution
Big Data for Social Good

Nuria Oliver: what big data and the Mexican pandemic taught us | Wired 2013 video

Impact of Government Measures on Mobility

0% 80% 55%

% Infected Agencies

Crime Prediction

70% accuracy

Socio-economic Level Inference

National Statistics Institutes carry out surveys on a subset of regions

Socio-economic status levels are inferred automatically for the rest of regions from Mobile phone data

Automatic inference of Socio-economic status is not monitored by the Government

82% accuracy for 3 levels

Analysis of impact of floods
Opportunities
Temporal and Spatial Granularity

Big Data can be available in real-time or if not in real time much more frequently than how data is typically collected (every 5-10 years for example for census data);
Some kinds of Big Data (e.g. data about the city, collected by sensors placed in the urban infrastructure) can be collected with significantly finer grained spatial granularity than with traditional methods;

Cost and Effort

Most of the Big Data that could be used for the public sector is data that has been collected already for other purposes. In addition, Big Data is typically collected by automatic means which makes its collection very cost-efficient;

Accuracy

It could be argued that some kinds of data that are relevant for official statistics (e.g. migrations) can be collected more accurately by automatic means through Big Data platforms than by manual means as it is the state of the art.
In addition, given that there isn’t a human-in-the-loop, the data is less prone to human errors and potential biases introduced by humans.
Challenges
Ebola and big data

Waiting on hold

Mobile-phone records would help combat the Ebola epidemic. But getting to look at them has proved hard

Oct 25th 2014 | From the print edition

IN THE battle against Ebola, mobile phones could be invaluable—not just in themselves, as devices that can be used to send people public-health information or let them call helplines, but also because of the data they generate. Phone companies use call-data records, or CDRs, to manage their networks and bill their customers. These records include a caller’s identity, the time of the call, the phone tower that handled it and the phone number of the other caller.
Internal Barriers

- Big Data for Development/Social Good projects are typically not part of any business unit in the Telcos.

- Therefore, these efforts are usually the result of a very motivated individual and/or small group of individuals within the company that champion these projects, many times in their personal time.

- Extracting appropriate datasets that are fully anonymized and aggregated, in compliance with local legislation and regulation, takes a lot of effort and resources. Unless there is strong internal support, it’s very hard to find the resources needed to do so.

- In order to accelerate, grow and eventually realize the potential of this area, it will be critical for the Telcos to recognize these efforts of strategic value to the company.
Risk/Benefit Analysis

• Even while adopting all sorts of precautions, if by any chance there is an error in the data extraction such that there could be a privacy leak, the consequences for the business could be devastating.
• Hence, the potential benefit would have to be really large to compensate for the risk.
• As this is an emergent research area, the benefits are still to be defined. We all know there is a lot of potential and we are carrying out case studies to better understand such potential, but it’s all still in a research stage.
Regulatory/Social

- Lack of updated regulation
- Lack of clear guidelines regarding safe data handling, processing and sharing for humanitarian purposes
- Risk of potential unintended consequences
- Risk of creating a digital divide, unbalanced access to data and-or expertise on how to analyze it and make sense of it
Technical

- Representativeness of the data, generalization
- Combination of data from multiple sources
- Real-time analysis and prediction
- Lack of ground truth → intervention to validate
- Significant vs substantially significant
- Correlations vs causality
Privacy/Security

• Potential privacy risks need to be minimized and understood. Control and transparency

• Security and traceability of the data

• Clear code of conduct and ethical principles when dealing with data

• Strict access control when appropriate
Framework for Data Sharing

Security
- Secure access to the data
- Anonymization

Exploratory
- Low to medium number of users

Developed
- Limited License
- Pre-computed Indicators & Synthetic Data

Remote Access
- Question and Answers

Applicative
- Medium to large number of users and open data

Limited License
- Data protected through

Pre-computed Indicators & Synthetic Data
- Exploratory

Anonymization
- Limited or aggregated data release

Authors:
- Yves-Alexandre de Montjoye (MIT, Harvard), Vincent Blondel (Louvain), Geoffrey Canright (Telenor), Nicolas de Cordes (Orange), Sébastien Delettaz (Real Impact Analytics), Kenth Engo-Monsen (Telenor), Sébastien Gambs (INRIA), Manuel Garcia-Herranz (UNICEF), Jake Kendall (Gates Foundation), Cameron Kerry (MIT Media Lab, Brookings, Sidley), Gautier Krings (Real Impact Analytics), Miguel Luengo-Oroz (UN Global Pulse), Nuria Oliver Ramírez (Telefonica), Zbigniew Smoreda (Orange Labs), Eric Wetter (Flowminder), Alex “Sandy” Pentland (MIT), Linus Bengtsson (Flowminder)
What can we all do to responsibly turn this opportunity into a reality?
Thanks!
nuria.oliver@telefonica.com
@nuriaoliver

Telefonica