



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.

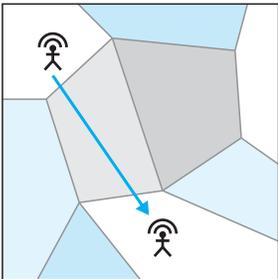
An additional piece of information that gets recorded in CDRs by a mobile network operator is to which cell towers the caller and recipient's phones were connected at the time of the call. Because the mobile network operator knows the locations of their cell towers, it is possible to use CDRs to approximate 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.

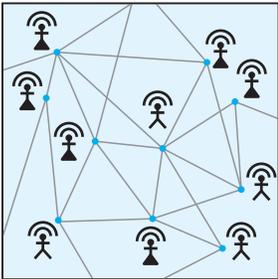
Internally, mobile phone companies use CDRs as the basis for billing customers and maintenance of their business, but CDRs can also serve other functions. Social scientists, researchers and public sector organizations have begun to research additional applications of CDRs. CDRs stored by a carrier have the potential to reveal personal information, so the records need to be altered in several important ways before being shared with third parties for analysis. First, all personally identifiable information must be removed. Typically, this is accomplished by encrypting the phone numbers of both caller and recipient. In many cases, the data from multiple callers may be aggregated to reduce risk of re-identification. Finally, data is often processed to contain the latitude and longitude from the cell tower closest to where the calls were placed. Consequently, by the time CDRs are shared, they look something like this:

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

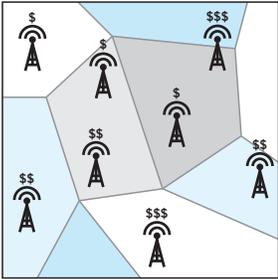
While at first glance it is difficult to assess the value of this rather rudimentary data, remarkably useful information on human behavior may be derived from large sets of de-identified CDRs. There are at least three dimensions that can be measured:



1. **MOBILITY:** As mobile phone users send and receive calls and messages through different cell towers, it is possible to “connect the dots” and reconstruct the movement patterns of a community. This information may be used to visualize daily rhythms of commuting to and from home, work, school, markets or clinics, but also has applications in modeling everything from the spread of disease to the movements of a disaster-affected population.



2. **SOCIAL INTERACTION:** The geographic distribution of one’s social connections may be useful both for building demographic profiles of aggregated call traffic and understanding changes in behavior. Studies have shown that men and women tend to use their phones differently, as do different age groups. Frequently making and receiving calls with contacts outside of one’s immediate community is correlated with higher socio-economic class.



3. **ECONOMIC ACTIVITY:** Mobile network operators use monthly airtime expenses to estimate the household income of anonymous subscribers in order to target appropriate services to them through advertising. When people in developing economies have more money to spend, they tend to spend a significant portion of it on topping off their mobile airtime credit. Monitoring airtime expenses for trends and sudden changes could prove useful for detecting the early impact of an economic crisis, as well as for measuring the impact of programmes designed to improve livelihoods.

CDR RESEARCH EXAMPLES

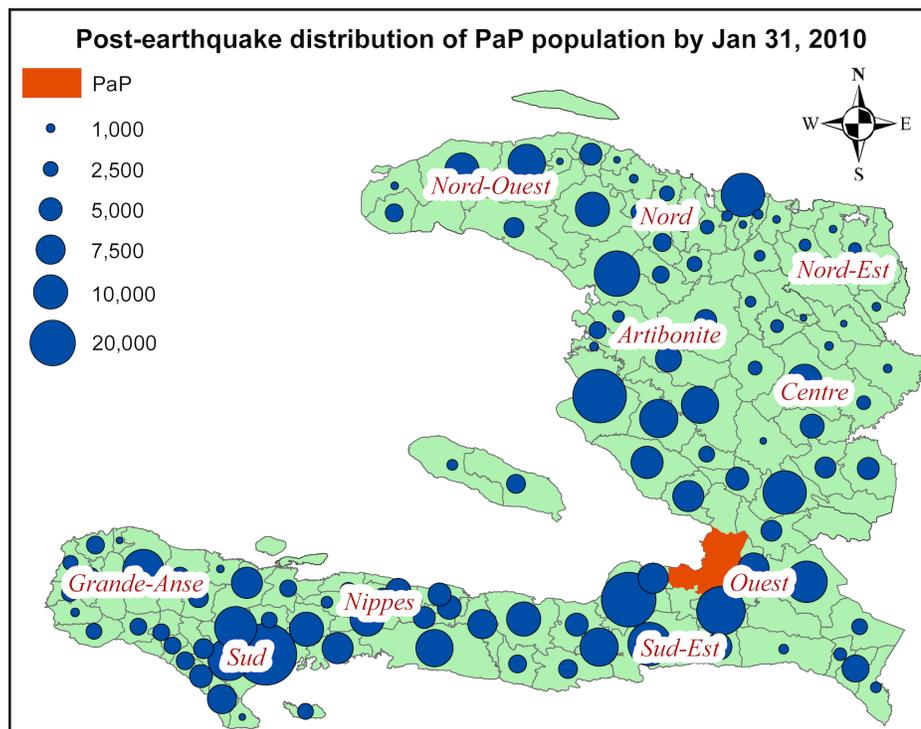
1. DISASTER RESPONSE

EMERGENCY MIGRATION IN HAITI

Natural disasters give rise to emergency situations where providing time-sensitive information is crucial for fast allocation of resources, which aids the response and recovery process. In 2010, a research team led by Sweden's Karolinska Institute showed that CDRs can be used to direct emergency aid by analyzing mobile phone records covering the time period when people are fleeing natural disasters (Bengtsson et al., 2011).

After the Haitian earthquake in 2010, many people moved away from the capital city, Port Au Prince. Researchers asked Digicell, Haiti's biggest mobile phone operator, to share de-identified information about the cell towers that subscribers were using when making calls. The data included the position of 1.9 million subscriber identity modules (SIMs) in Haiti from 42 days before the earthquake to 158 days afterwards, allowing researchers to compare people's movement in the days preceding and following the earthquake. This study proved more accurate in measuring the number of displaced people and their destinations than the studies by the Haitian Civil Protection Agency, and the estimates of geographical distribution of people across Haiti were matched by estimates from a retrospective United Nations Population Fund study.

A follow-up study (Lu et al., 2012) found that the destinations of people who left the capital during the first three weeks after the earthquake were correlated with the locations to which they had significant social bonds. This research has shown that in the event of another natural disaster, population movement patterns may be significantly more predictable than has been previously understood. This type of analysis could be used to plan relief efforts more precisely.



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

Figure 1: The visualization shows the distribution of population migration from Port au Prince (PaP) after the Haiti earthquake obtained by analyzing CDRs. The circles represent locations that received at least 500 people from the estimated distribution of those in PaP on the day of the earthquake, but outside the city 19 days after the earthquake. Source: Lu et al., 2012.

2. HEALTH

MAPPING MALARIA IN KENYA

Malaria kills about one million people each year. In sub-Saharan Africa, 90% of malaria-related deaths are of children under five years old (Harvard School of Public Health, 2012). When an Anopheles mosquito bites an infected person, a small amount of blood infected with microscopic malaria parasites is withdrawn. When the mosquito bites someone else, the parasite is injected with the bite, and the transmission of malaria is complete, resulting in further spread of malaria. Therefore, to predict malaria's geographic spread, it is important to factor in not only information about the location of the mosquitoes that carry the malaria parasite, but also the behavior of people who might be infected.

Researchers analyzed mobile phone data to study the regional travel patterns of nearly 15 million mobile phone subscribers in Kenya, over the course of a year (Wesolowski et al., 2012). They estimated the daily locations of 14.8 million Kenyan mobile phone subscribers between June 2008 and June 2009, mapping calls and texts to one of the towers located within the boundaries of 692 settlements. The researchers analyzed the mobile phone data together with a simple malaria transmission model based on infection prevalence data, and in doing so were able to map routes of malaria parasite dispersal. The study included people who became infected after residing in areas where the disease broke out, and those who visited areas with outbreaks, became infected and returned home. After analyzing the CDR data and using the transmission model, scientists were able to map the specific locations where malaria outbreaks originated, and the locations where the disease had a higher probability of spreading. They found that a large proportion of “imported” infections—for example, infections that are carried by people moving from one place to another—end up in Nairobi, with infected residents returning there after journeys to hot spots such as Lake Victoria.

The study indicates that assessing connectivity among different regions of Kenya can help estimate costs for regional elimination strategies, identify ‘source’ regions and pinpoint anticipated hot spots. These hot spots can then be targeted with use of insecticides, drugs and bed-nets and with measures such as reducing stagnant water and destruction of old tires. On a broader scale, this analysis can be used to identify the high-volume human traffic between regions that increases transmission rates, and then communicate risks to travelers to alter their behaviors, restrict travel and employ surveillance mechanisms in high-risk areas.

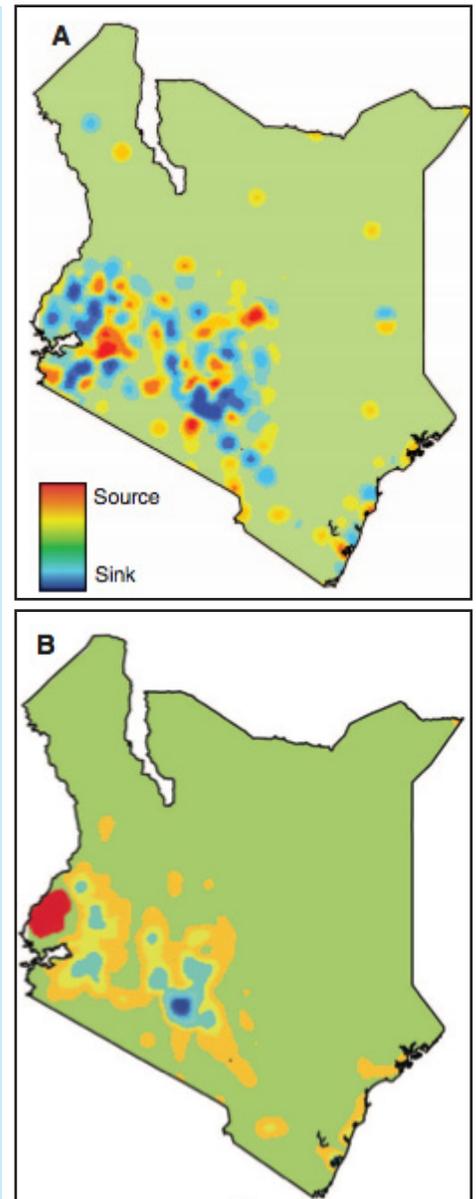


Figure 2: Figure shows sources and sinks of human travel and parasites. Kernel density maps show ranked sources (red) and sinks (blue) of human travel (2A) and total parasite movement (2B) in Kenya, where each settlement was designated as a relative source or sink based on yearly estimates. Source: Wesolowski et al., 2012.

By analyzing the regional travel patterns of millions of mobile subscribers, researchers were able to map the specific locations to which malaria had a higher probability of spreading.

COMBATING H1N1 FLU IN MEXICO

To control the spread of the H1N1 flu epidemic in 2009, the Mexican government imposed mobility restrictions on citizens. Understanding and quantifying the effect of such restrictions had not been adequately studied on a large scale before. Researchers measured the impact of the policies taken by the government to limit human mobility, using CDRs from one million anonymized customers in one of the most H1N1-affected Mexican states (Frias-Martinez et al., 2012). The data covered a period of January to May 2009. The government issued alerts in three stages: a medical alert (stage one), closing schools and universities (stage two) and suspension of all non-essential activities (stage three). It was found that up to 80% of the population reduced their mobility during stages two and three. However, the analysis also revealed an increase in the number of visitors to the airport before the suspension of all non-essential activities (stage three). This behavior might have limited the containment causing an undesired spread of the epidemic. Despite this, the results indicated that the mandates managed to reduce the peak number of individuals infected by the virus and postponed the peak of the pandemic by two days, allowing authorities to react faster to control the epidemic.

Quantitative analysis of CDR data allowed researchers to measure the effectiveness of government mandates on people's mobility during H1N1 outbreak in Mexico.



H1N1

H1N1, or swine flu, is a potentially fatal subtype of the human influenza virus. H1N1 was the source of a worldwide pandemic in 2009, including in the United States and Mexico.

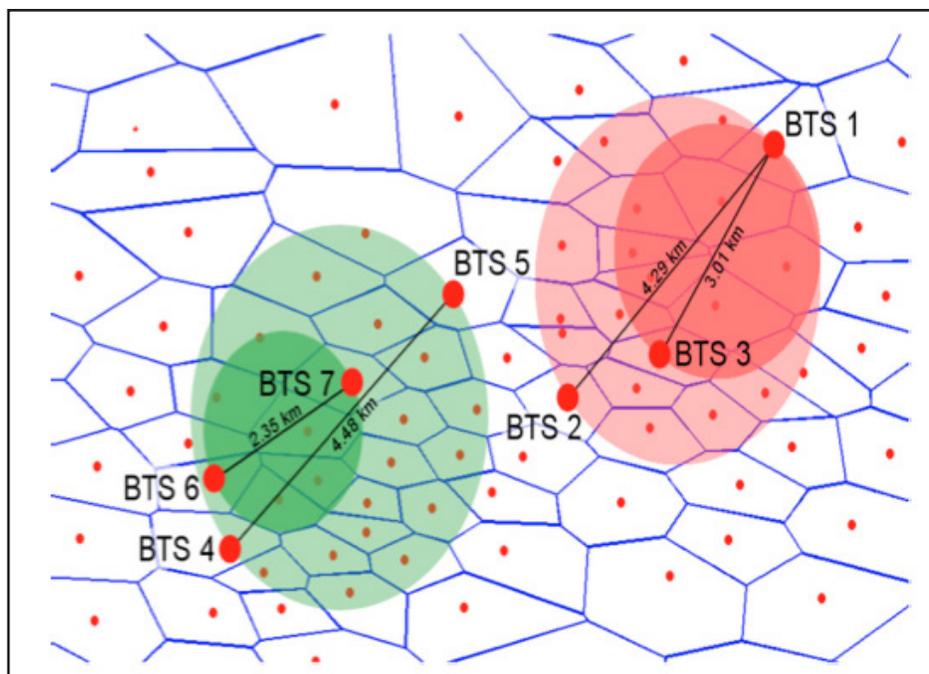


Figure 3: The visual shows the geographical region where the daily activities of an individual take place—the area of influence—for two individuals (one marked in green and the other in red). The bigger circles represent the mobility behavior under normal circumstances and the smaller circles represent the mobility behavior under the epidemic alert period. The red dots represent mobile tower stations (labeled BTS). In the case of the region marked with red, the baseline diameter of mobility is 4.29km (defined by BTS1 and BTS2). During the alert period, the diameter is reduced to 3.01km (the distance between BTS1 and BTS3). For the region marked in green, the baseline period has a diameter of 4.48km (defined by BTS4 and BTS5), and for the alert period the diameter is reduced to 2.35km (defined by BTS6 and BTS7). Source: Frias-Martinez et al., 2012.

3. SOCIO-ECONOMICS

UNDERSTANDING SOCIO-ECONOMIC INDICATORS IN THE UK

Prior studies have shown that economic opportunities are likely to come from contacts outside of tightly-knit local social groups. However, the correlation between network diversity and a population's economic well-being had not previously been quantified. To better understand this relationship, researchers studied CDRs for August 2005 in the UK (Eagle et al., 2010). The data contained over 90% of the country's mobile phones and almost 100% of the residential and business landlines. Researchers coupled this data with census information on socio-economic well-being from the UK Government's Index of Multiple Deprivation. The findings revealed that the diversity of individuals' relationships is strongly correlated with the economic development of communities. Researchers were able to validate a central assumption that was untested at the population level: that more diverse ties correlate with better access to social and economic opportunities. Although further research is needed, this valuable insight is particularly relevant to policymakers engaged in targeted poverty reduction initiatives.

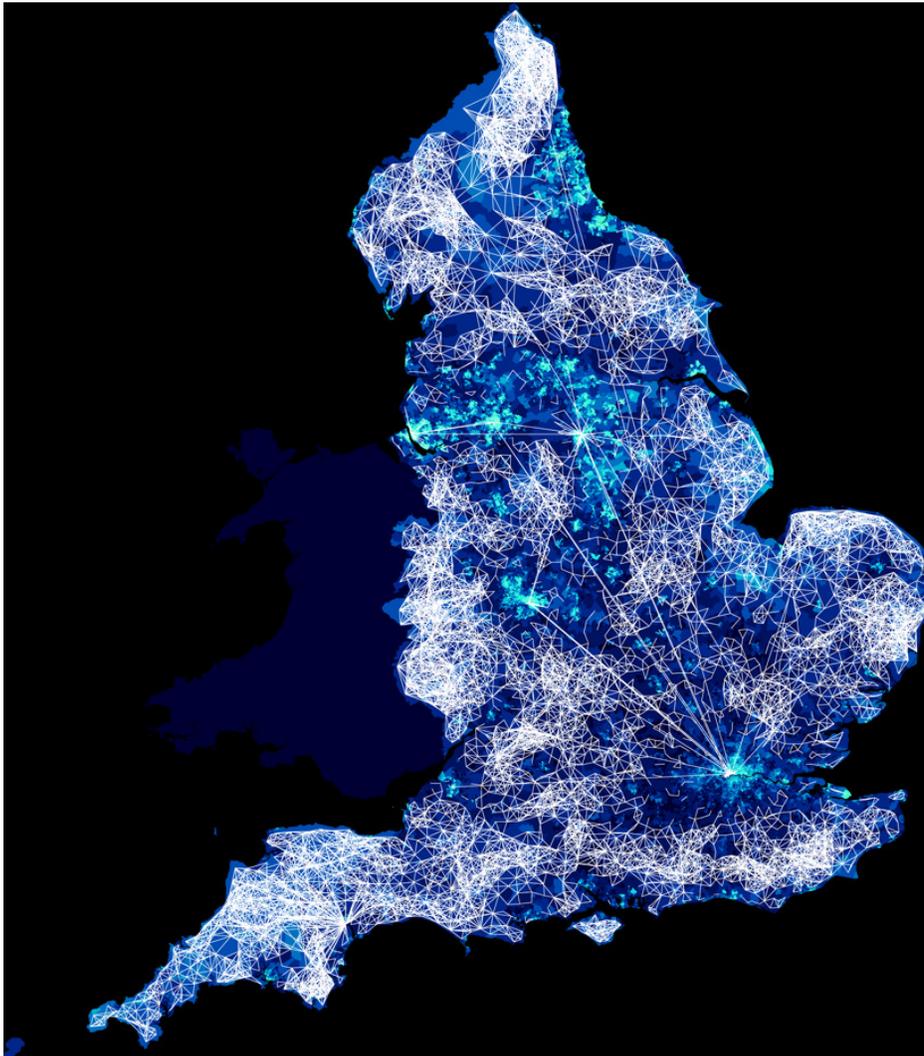


Figure 4: This is an image of regional communication diversity and socioeconomic ranking for England. Communities with diverse communication patterns were found to rank higher (represented from light blue to dark blue) than the regions with more insular communication. This implies that communication diversity is a key indicator of an economically healthy community. Source: Eagle et al., 2010.

PROVIDING PROXY CENSUS MAPS IN LATIN AMERICA

This study (Frias-Martinez V, Virsesa, 2012), like the previous example regarding socio-economics in the UK (Eagle et al., 2010), focuses on socio-economic factors and their relationship to cell phone usage in Latin America. Researchers at Telefonica Research in Spain studied whether specific gender, age or socio-economic groups use mobile phones in different ways, using a new analytical approach that combined CDRs with countrywide census data gathered by the National Statistical Institute. The data set contained five months of CDRs from over ten million subscribers across twelve cities. Results showed a strong correlation between the socio-economic level of a person and the expenses, reciprocity of communications, physical distance with his/her contacts and the geographical areas around which people move. The study was also able to provide a method to approximate socio-economic indicators from calling patterns, which could provide an affordable tool in the context of resource-constrained economies.

ESTIMATING POVERTY LEVELS IN CÔTE D'IVOIRE

In another study (Smith et al., 2012), researchers used CDRs to map poverty levels in Côte d'Ivoire. No full survey of the country's population has been published since a civil war in the 1990s. Researchers used anonymized CDRs of five million Orange telecommunications customers between December 2011 and April 2012 to assess both the level of activity among subscribers and locations where calls were made. Higher levels of mobile communication and wider range of calls are a proxy indicator for prosperity. Using this data, poverty levels of eleven regions of Côte d'Ivoire were quantified. The estimate was validated when compared with a multi-dimensional poverty index created by University of Oxford, which uses indicators such as poor health, lack of education, inadequate living standard and threat from violence among other factors. This research validated the possibility of making poverty maps using CDRs.

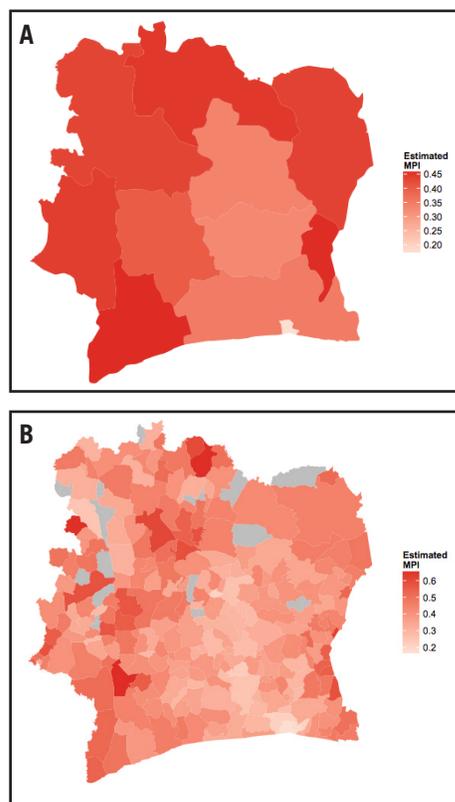


Figure 5. Figure 5A shows poverty map estimated based on the antennas in the eleven major regions of Cote d'Ivoire, where the darker areas indicate higher estimated poverty level. Figure 5B shows the Department poverty levels as approximated by the model used on regional level indicating the finer granularity possible when using CDRs. Source: Smith et al., 2012.

The previous three studies show that in countries where information regarding census, socioeconomic levels or poverty indicators is lacking, CDRs may be used to estimate these variables on an ongoing basis, thus augmenting and complementing survey data.

CDRs can provide a proxy indicator for assessing regional poverty levels, and can valuably augment national surveys in estimating changes associated with a growing economy.

4. TRANSPORTATION

OPTIMIZING TRANSPORT NETWORKS IN ABIDJAN

Rapid urbanization in developing countries has increased pressure on infrastructure such as road networks. Roads and public transportation systems become saturated, and people lose a great deal of time traveling from home to work, which in turn has a collective cost on the burgeoning economies. Researchers at IBM's AllAboard project have shown it is possible to monitor citizens' travel routes and use data-driven insights to better plan and manage transportation services (Berlingerio et al., 2013). The data analyzed included CDRs shared by Orange for 500,000 users over a five month period, which were used to pinpoint locations based on cell towers used when making calls or using a mobile-based service. The researchers studied CDR data, mapped against 85 bus routes in Abidjan, Ivory Coast's largest city, where the bus transport network spans 539 buses, 5,000 mini-buses and 11,000 shared taxis. Their findings allowed the researchers to suggest a partial solution to the city's congestion: add four routes to Abidjan's existing infrastructure and extend another route. This optimized network would reduce travel time by 10%. Such a provisioning method using CDRs would be useful for better urban planning and public transportation.

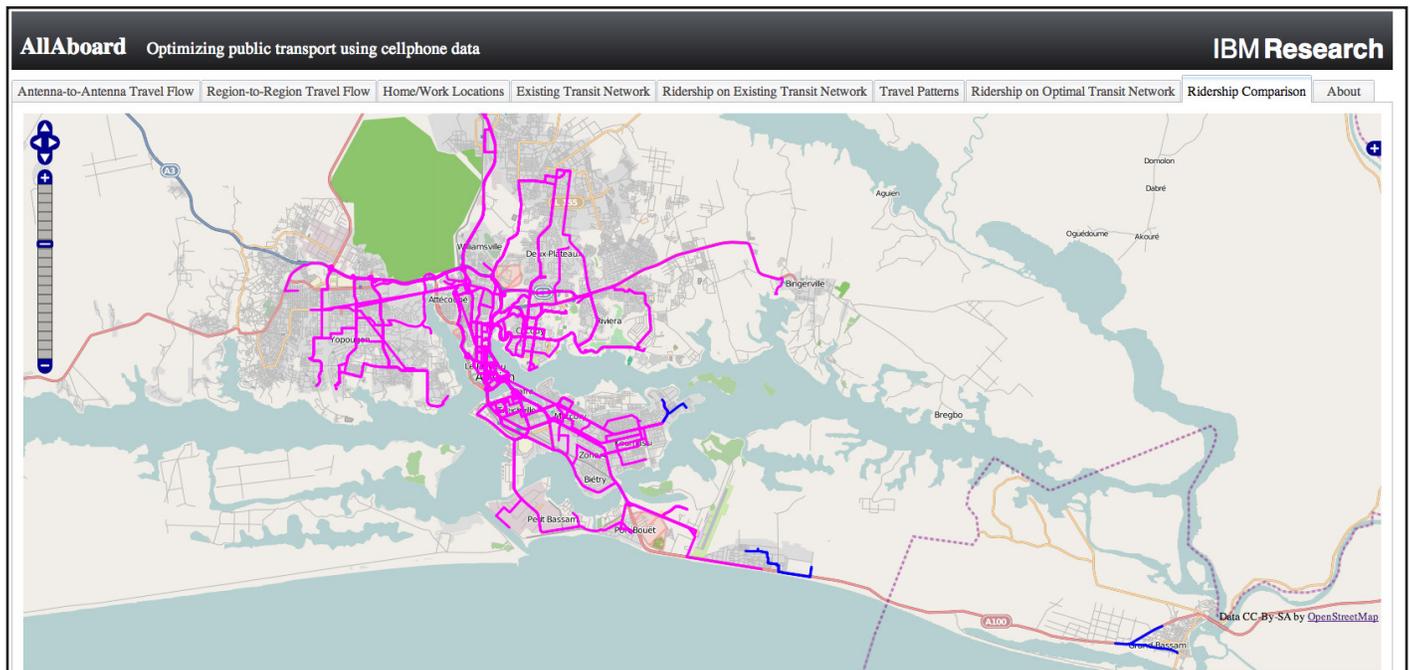


Figure 6: This image shows the existing public transport network (SOTRA) in Abidjan and additional routes suggested by the study. Source: Berlingerio et al., 2013.

By analyzing CDR data, scientists mapped new routes to decongest Abidjan's crowded roads, which would reduce travel time by 10%.

PRIVACY CHALLENGES

There has yet to be a commonly accepted definition of the term “privacy.” As defined by OECD, privacy is “an individual’s freedom from excessive intrusion in the quest for information and an individual’s ability to choose the extent and circumstances under which his or her beliefs, behaviours, opinions and attitudes will be shared with or withheld from others” (OECD, 2005). The majority of risks associated with the misuse of data concern *personal data* or *personally identifiable information*. For this reason, anonymization is common practice when conducting Big Data research.

Robust anonymization of data has been recognized as a critical component of successful research, and academia conforms to strict institutional standards for ethical use of subject data. However, there is increasing awareness regarding the possibility of re-identification of anonymized data. In the Big Data research field, *linkability* refers to the cross-referencing of multiple anonymized data sets, which can be used to recover the identities of people whose data has been anonymized.

It remains true that the safest way to preserve privacy is to irreversibly anonymize and aggregate data sets, but experts continue to explore how to preserve privacy while keeping data sets granular enough to yield valuable insights for public good or social science research.

Several studies have analyzed the exact threshold at which individuals become identifiable. Director of the Harvard University Data Privacy Lab, Dr. Latanya Sweeney, introduced the concept of *k-anonymity*, a way to quantify the order of the risk of personal identification from a data set which decreases in magnitude as the data is made artificially less granular (Sweeney, 2002).

Another recent project used mobility data from 1.5 million mobile phone subscribers in a small Western country, where the location of an individual was specified hourly with a spatial resolution equal to that given by the carrier’s antennae in a small Western country, to demonstrate that just four spatio-temporal points were enough to show uniqueness of 95% of the individuals (de Montjoye et al., 2013). By understanding the privacy bounds of human mobility, the data can be ‘coarsened,’ or made less specific, in order to minimize the chance of personal identification.

Data linkability has also been heavily considered in the health sector. Studying the risks of re-identification, Professor Khaled El Emam offered a system that determines to what degree data sets should be de-identified before release (El Emam, 2010). He suggests the risk can be quantified as a *maximum* level—which is equivalent to the concept of *k-anonymity*—and quantified as an *average* level, the latter being more appropriate for non-public data releases (El Emam, 2013).

Refraining from sharing one’s personal information, even when tightly controlled and in an anonymised or aggregated form, is clearly appealing on an individual level. However, professor and privacy lawyer Jane Yakowitz examines privacy risks versus the value of Big Data for research, and argues that this comes at the greater cost of losing significant policy insights which rely on a large representative sample of data (Yakowitz, 2011).

Ethical guidelines, legal frameworks and technological solutions for protected data sharing are maturing in parallel with the evolution of mobile phone usage around the world and the increased use of Big Data for development and humanitarian purposes. When privacy principles are at the center of the process in any research or technological development, it is referred to as “privacy by design.” Proceeding in this way promotes compliance with existing ethical standards by implementing privacy protective mechanisms from the beginning.



PRIVACY BY DESIGN

Privacy by design is a practice that embeds privacy and data protection principles in all research or technological development and across the entire project life cycle.

BIBLIOGRAPHY

DISASTER RESPONSE

Bengtsson L, Xin Lu, Thorson A, Garfield R, von Schreeb J (2011) Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti, *PLoS Medicine* 8(8).

Lu X, Bengtsson L, Holme P (2012) Predictability of population displacement after the 2010 Haiti earthquake, *Proceedings of the National Academy of Sciences* 108(29) 11472-11473.

HEALTH

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.

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.

SOCIO-ECONOMICS

Eagle N, Macy M, Claxton R (2010) Network diversity and economic development, *Science* 328(5981) 1029-1031.

Frias-Martinez V, Virsesa J (2012) On the relationship between socio-economic factors and cell phone usage, *International Conference on Information and Communication Technologies and Development*.

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*.

TRANSPORTATION

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*.

PRIVACY CHALLENGES

OECD (2005) Glossary of Statistical Terms. Retrieved from <http://stats.oecd.org/glossary/detail.asp?ID=6959>

de Montjoye Y, Hidalgo C, Verleysen M, Blondel V (2013) Unique in the crowd: the privacy bounds of human mobility, *Nature Scientific Reports* (3) 1673. Retrieved from <http://www.nature.com/srep/2013/130325/srep01376/full/srep01376.html>

Sweeney L (2002) k-Anonymity: a model for protecting privacy, *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems* (10) 557-570.

El Emam K (2010) Risk-based de-identification of health data, *Security & Privacy, IEEE* 8(3) 64,67. doi: 10.1109/MSP.2010.103

El Emam K (2013) Guide to the de-identification of personal health information. CRC Press (Auerbach).

Yakowitz J (2011) Tragedy of the data commons, *Harvard Journal of Law & Technology* (25) 1.

ACKNOWLEDGMENTS

This document has been drafted, edited and produced by the Global Pulse team: Sara Cornish, Olivia De Backer, Anirvan Ghosh, Shehryar Kayani, Robert Kirkpatrick, Jonggun Lee, Tomaz Logar, Miguel Luengo-Oroz, René Clausen Nielsen, Jennifer Poole, Liudmyla Romanoff, Alex Rutherford, Anoush Tatevossian, Makena Walker.

We would like to thank Robert Bond, Khaled El Emam, Kenth Engø-Monsen, Vanessa Frias-Martinez, Yves-Alexandre de Montjoye, Jules Polonetsky and Jane Yakowitz for reviewing drafts and providing feedback on this document.

Contact: info@unglobalpulse.org.



ABOUT GLOBAL PULSE

Global Pulse is an innovation initiative of the United Nations Secretary-General exploring how new, digital data sources and real-time analytics technologies can help policymakers gain a better understanding of changes in human well-being and emerging vulnerabilities. Through strategic public-private partnerships, innovative analysis and open source technology development across its network of Pulse Labs, Global Pulse develops approaches for applying Big Data to 21st century development challenges.

The three-fold implementation strategy includes:

- 1. Research & Development:** Conducting research to discover new proxy indicators for tracking development progress and emerging vulnerabilities in real time, and assembling a toolkit of technologies for analyzing real-time data.
- 2. Big Data partnerships:** Forging partnerships with companies, organizations, researchers and academic institutions that have the data, technology and analytical expertise needed for Big Data for Development projects and advocacy.
- 3. Pulse Lab network:** Establishing an integrated network of country-level innovation centers that bring together government experts, UN agencies, academia and the private sector to prototype and pilot approaches at country level.

For more information please visit www.unglobalpulse.org.