Leveraging mobile data for sustainable development

Last updated September 20, 2016

The objective of this module is to describe an approach for leveraging mobile data to address social goals and questions in support of the SDGs. To support more real-time, dynamic and disaggregated data for decision making and fill data gaps where possible, satellite and mobile data can be immensely important. While generally, emerging countries may have more limited insights on their population through large-scale integrated data sources, the application of mobile data to support these types of analysis is universal. Mobile data show multiple advantages such as covering large portions of the population, offering standard format and providing insights on both individual and social behaviors.

For the Data for Action section of this Toolbox, mobile data represent an innovative, cost-effective and dynamic source of data to support the SDGs considering the scale of information available – nearly 5 billion people around the world own a mobile phone. The potential is tremendous. However, this is still very much a developing field especially in regards to the privacy and regulation, institutional and legal, and governance considerations. As such, this module is being developed iteratively and this initial release concentrates more on the what and how explaining mobile data, its potential uses, and some business models for how it can be accessed. Further iterations will dive deeper into these considerations, as well as refine the use cases represented in this paper.

Table of Contents

The need: wider, deeper and more useful data—that contribute to the social good

There are now nearly 5 billion mobile phone subscribers globally—which means an unprecedented amount of user data exist that can be leveraged for a multitude of purposes. The proliferation of digital services—from mobile operator call data records, to mobile apps digital services, to value added services (VAS) delivering health, education and other facilities over mobile channels—holds the potential to generate unprecedented volumes of data.

In parallel, new techniques and business models for drawing insights from and analyzing multiple and large datasets are becoming more commonplace. The scale of use of mobile phones and other digital services is enabling data scientists to determine location and movement patterns, financial and economic activity, identity and demographic data, click/browser activities, social graphs and call patterns, sentiment and trends and even diagnostic and ambient conditions surrounding a user. This is enabling new ways of understanding user needs for both private commercial and public good uses.

In many low- and middle-income countries, knowledge on the distribution of the population and its behaviors is less precise and granular, and the ease or even possibility of measurement can be very limited. It is thus difficult to effectively target projects and actions to reduce poverty and improve well-being. As nearly 70% of the bottom fifth of the population in developing countries now own a mobile phone—more than the proportion who have access to electricity or clean water—we are now in a position where a huge amount of mobile data could be used to develop social good programs at scale.

Interest in the use of these large sets of digital data for social impact is growing, particularly as the development community overall is increasingly paying attention to the use of data of all kinds for decision-making. In addition to using such data to refine service delivery to better meet user needs and to enable economic sustainability for low-cost digital services, a wide range of use cases are possible, including in epidemiology, migration tracking, financial product development and more. Mobile data could thus be immensely important to support fact-based decision-making and hence the implementation of programs around poverty reduction and other social goals.

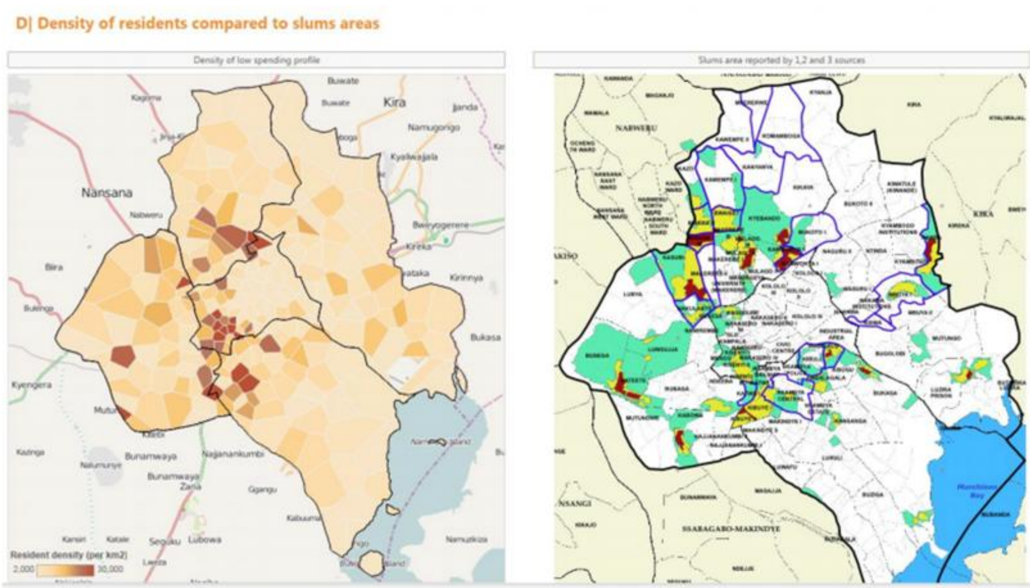
Mobile data reflect particular telecom operations: the market is partial and not stable, entails a wealth and gender bias and can sometimes generate spurious correlations. But, if we understand the bias, the data can also be faster, cheaper and more granular and can answer to completely new indicators. While keeping expectations in line with reality, we can aim high to fulfill the potential that mobile data offer.

The current aim: a framework for unlocking mobile data for social uses

This document aims to introduce ways to leverage mobile data for use in anonymous formats to support measurement, reporting and decision-making for the social good. It offers a framework and guidelines for practitioners and local stakeholders in both data-rich and data-poor environments, and also presents a list of key elements to secure in ensuring the monitoring, tracking and prioritizing of the global Sustainable Development Goals (SDGs) within the national or regional digital context.

In real terms, this document first looks at data we can access and under what business models, and then the very practical things we can do by leveraging such data. This is the '*what*' in the approach. It is here that it will, at least initially, be most beneficial: presenting examples of what we can actually *do* with the data. All actors, including not only those funding and implementing sustainable development programs on the ground, such as donors, country governments, non-governmental organizations, but also industry stakeholders, national statistics organizations and the local private sector, will be able to access information on the value of the metadata, to see how they can apply them within their own work, and in particular for social good within policy-making and planning.

There are also many questions about '*how*' to leverage these mobile data for use in this way. This document is a work in progress: its second part will detail the complexities involved in operationalizing this approach, and the precautions to bear in mind, with the aim being to analyze these questions in more depth later on as the discussion evolves.



Source: Real Impact Analytics (2016) Leveraging Data to Address Untapped Social Dimensions.

What are the data?[1]

"Mobile data" here are defined as the metadata generated by the individual users of telecom operator services. The customer's mobile device provides the first set of data, through call data records (CDRs) and, for example, recharge information, mobile money logs and mobile internet data: personal information, contacts, social graph, basic clickstream data and user location and movement patterns. As people's mobile devices become "smarter", researchers can obtain richer insights—such as accurate location and information on diagnostic and ambient conditions. For the most part, any data obtained directly from the user's device would require users to "opt in" and grant consent to access.

Mobile phone networks transmit signals carrying voice, text and digital data through thousands of overlapping geographic areas, or cells, each with its own base transceiver station (BTS). The BTS is part of the base station subsystem (BSS), which is responsible for handling traffic and signaling between a mobile phone and the network. The base station controller (BSC) controls the intelligence that manages multiple BTSs and the traffic flow between them, as well as other core functions. In general, base station (or tower) data can provide interesting insights into user location and movement, though this information is available only to the operators for internal network planning or to law enforcement.

Radio frequency signals are transmitted by a mobile device and received by the base station. They travel through backhaul, which is the intermediate wireless communication infrastructure that connects smaller networks with the backbone or the primary network, using microwave, fiber optic or copper. At this stage, signal characteristics such as attenuation and signal distortion can be measured to provide an indication of local ecology, rainfall patterns, civil construction, etc.

In time, current CDRs will disappear and use of mobile internet data will increase, for example through Skype and Whatsapp. This means we are going to lose richness in the data: we won't be able to see whom a caller is actually calling, although we will be able to track the actual making of that call on, say, Skype. This will make it necessary to go in more depth into privacy issues in the future: corporations such as Google, Facebook and LinkedIn will be able to gain more insights into customers and populations but we will not have access to these data. We will need to be more flexible and adaptable in the future, and to work with the situation as it evolves.

These mobile data have multiple advantages: they cover large portions of the population in a standard format, and can be used to specify location and within a host of other prospective areas to generate a "behavioral signature", which can be used to generate predictions on human activity. Complementary data sources will be needed in all cases, for triangulation and for rigor. We will also need actual in-country scientific research to accompany the leveraging of the data, to identify any correlation between the data and the actual reality in a country. Machine learning enables us to use the data, matched with relatively small-scale surveys, to train an algorithm to see more precisely how humans are behaving in a given area and therefore what their needs might be.

[1] This section draws on Naef, E., Muelbert, P., Raza, S., Frederick, R., Kendall, J. and Gupta, N. (2014) *Using mobile data for development*. Report by Cartesian for Financial Services for the Poor Team at the Bill & Melinda Gates Foundation.

Business models

Developing algorithms or tools involves accessing data to design the underlying code, to test the outcome and to operationalize the end product. There is hence a technical need to access the data regardless of the underlying business model. The data services value chain rest on four key steps:

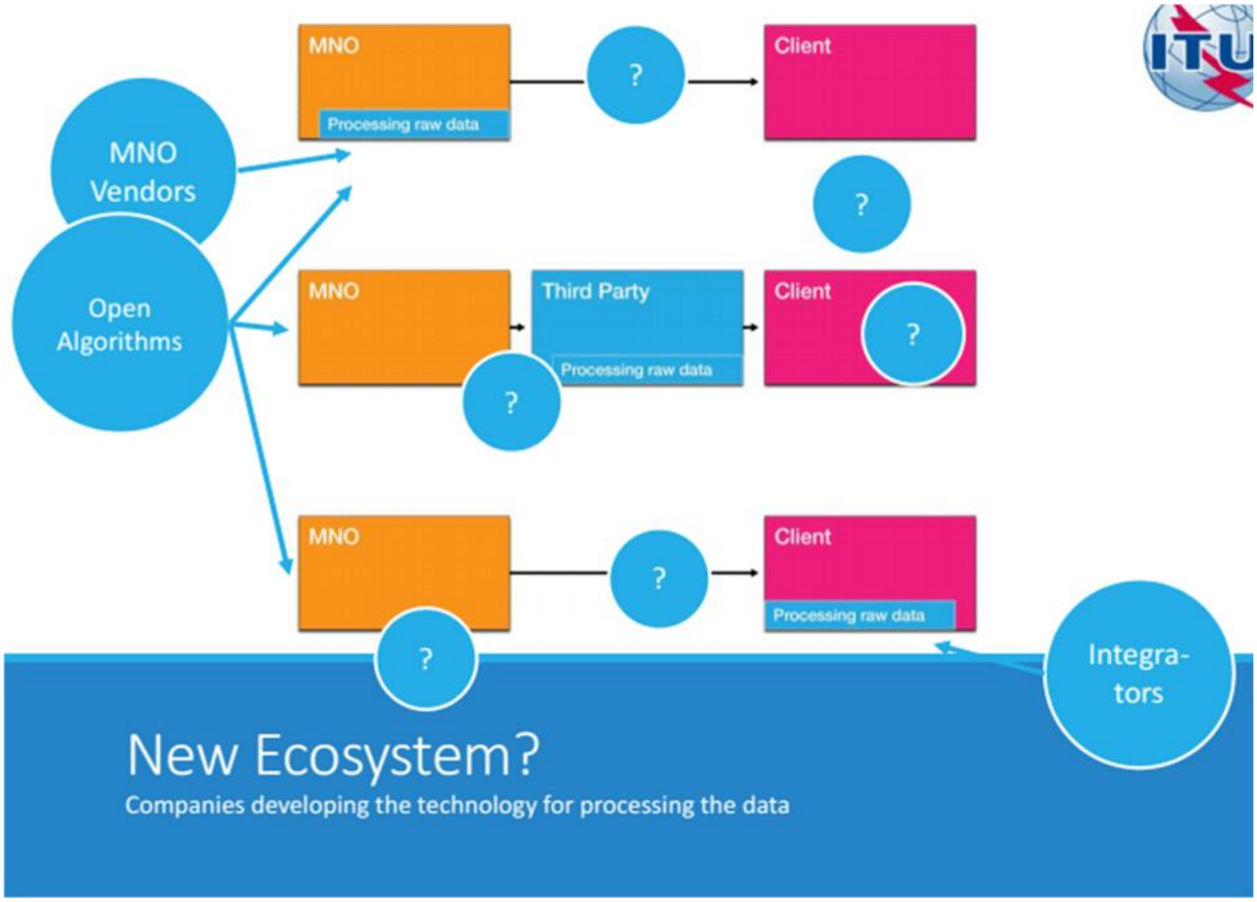
1. Data collection.
2. Data preparation and storage, with some preparation involved.
3. Data analytics and data visualization (the vertical added value analytic services).
4. Service distribution (consulting, apps, trainings, etc.).

To support this, there exist multiple business models; there is no need to pick any at such an early stage. A mechanism that entails open finance would be preferable, to allow for the transformation of the data for greater use globally. We have identified essentially five models:

- In-house production of statistics: Data are stored by the telecom operators and they produce aggregates and statistics internally
- Trusted third party: both the data owner and the data users cannot support the security burden of hosting the data themselves, or wish to blend this data with other data sources (with the individuals' consent), but can't do this themselves for obvious privacy reasons. In this case, they can turn to a trusted third party that will host the data themselves, and provide services to enable a secured access to this data.
- Open data model: Data are available for any registered developer. They might be open at different levels of aggregation. Developers and end users design and operate their own algorithms and tools based on the available open data.
- Open algorithm model: Telecom operators or third-party providers develop open algorithms based on proprietary data. Developers and end users freely use the open algorithms within their own tools and need to connect their own data sources. Algorithms are necessary across platforms, which is more complex but still feasible.
- Marketplace model: Third-party providers develop algorithms and tools based on proprietary data. Developers and end users remunerate third-party providers to feed their tools with mobile data and to operate and maintain these tools.

Multiple combinations of the above models could exist:

- Mobile metadata are available to one third-party provider.
- One third-party provider offers access to aggregated data on a platform hosted in a regional cloud.
- Other third-party providers develop and operate their own algorithms and tools and access aggregates available on the regional cloud.
- Some other third-party providers develop open algorithms and tools.
- Any end user pays a third-party provider to (i) access mobile data and remunerate the telecom operator; and (ii) use their algorithms and tools or freely use open algorithms.



Source: Tiru, M. Business Models for Mobile Big Data.

What we can use the data for: social and commercial action and decision-making

We are now working on unlocking the potential of this private data to support more effective action and decision-making globally. Mobile data are offering us the opportunity to generate new indicators and proxy indicators that are easier to capture, more agile, cheaper and more up-to-date and yet have similar fidelity to time-tested ways of carrying out measurement and thus contributing to decision-making. This will be vital both in the commercial field and for social development programming.

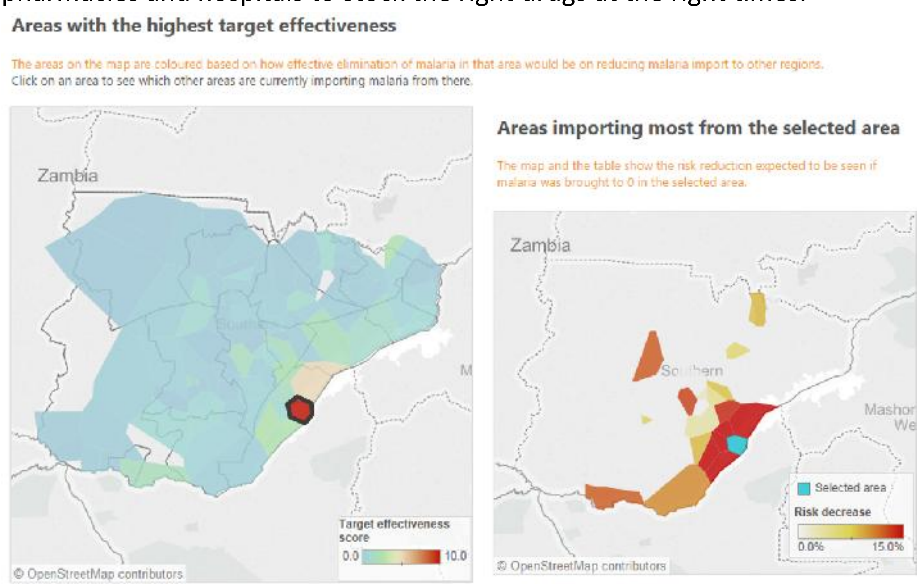
We will continue to conduct machine learning in emerging countries to train algorithms to predict human behavior on different socio-economic indicators and proxy indicators. These predictions will represent faster, more reliable and more precise contributions into improving business conditions and also ensuring better performance on poverty reduction and social well-being.

Location-based information is the most direct: metadata can be used to take a census of the population, look at commuting patterns and analyze migration and long-term inter-city mobility. Other variables within the metadata may be able to help us work in more prospective areas, on issues such as socio-economic behaviors. In terms of proxy indicators, for instance, mobility can be measured and proxied to generate information on other areas, such as in health, education or poverty, and in itself represents an essential element in national decision-making on, for example, infrastructure planning.

The aim is to leverage mobile data to measure those areas where there is limited or a total absence of measurement currently, to fill in the gaps in our knowledge the most efficiently.

A few examples of what we can do with mobile data are as follows (see Appendix for a framework of possible use cases aligned with the SDGs):

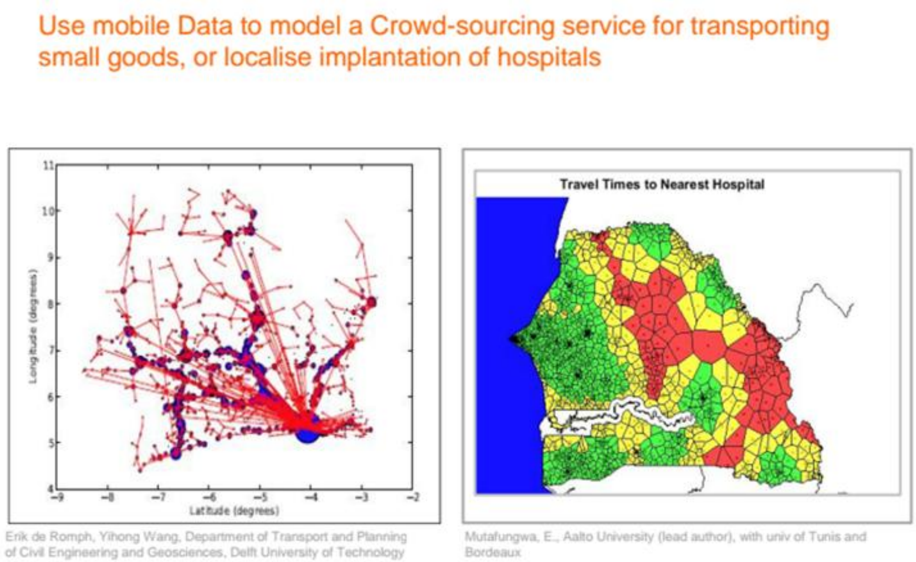
- **Food security.** Purchase of mobile phone top-ups[2] can be correlated with food security: if a person has just sold their harvest, they may be able to afford to feed their phone—as well as themselves. Lower purchase of top-ups may correlate with lower food spending in market-dependent households, and thus indicate a situation of lower food security, and at least a demand for human interpretation of the need for intervention.
- **Health.** Data on, for example, mobility around health centers or pharmacies, correlated with data on pharmacy sales, can be proxied to highlight disease occurrence and develop epidemiology models. Mobility data show us who is moving where and in what numbers, helping us calculate a disease vector and thus predict how it will spread. The analysis can then be triangulated to help send vaccines and doctors with more precision, and also enable pharmacies and hospitals to stock the right drugs at the right times.



Source: Data-for-Good @ Real Impact Analytics (2016)

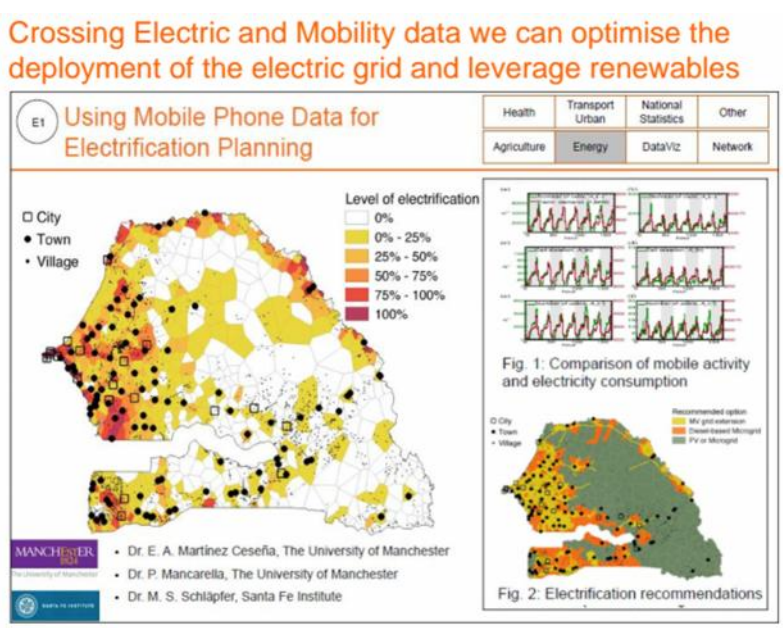
---

[2] Top-ups are a means of adding and storing value in a mobile phone in pre-paid environments.



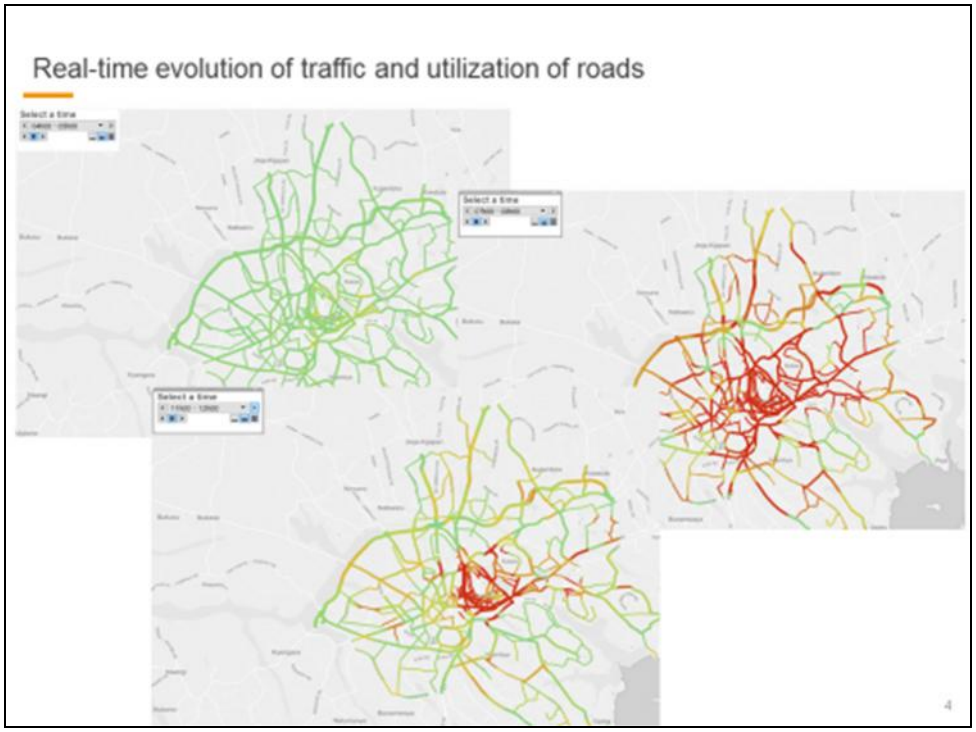
Source: de Cordes, N. (2016) Data for Development and OPAL.

- **Water.** On certain antennas, if there is a drought, it will be possible to see a great deal of phone activity around wells and water sources. Although there may be reasons other than drought for such activity, it will now be possible to raise a query and interpret the location and mobility data in a human way to ensure interventions are possible if necessary.
- **Energy.** Along with measuring the percentage of the population with electricity access, we can develop a model for optimal network deployment using the correlation between mobile and energy usage. We can also help reduce the cost and increase the quality of electricity by feeding real-time load management.



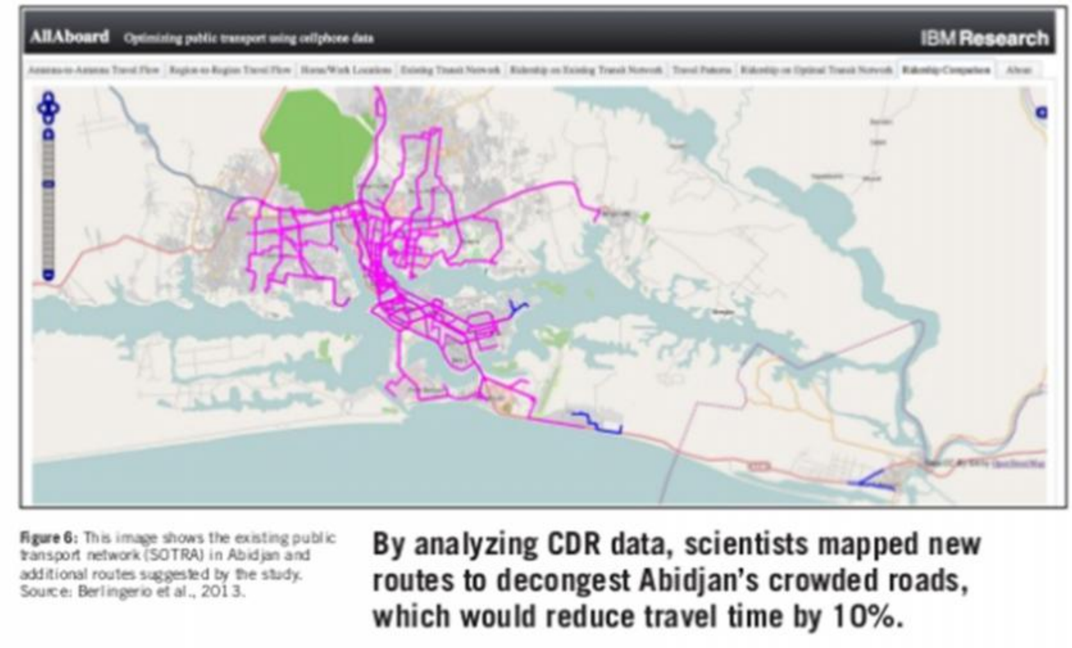
Source: de Cordes, N. (2016) Data for Development and OPAL.

- **Infrastructure.** Mobility data can be used to analyze the need for large investments in infrastructure, for example in dams, electrification, roads and railways. The data can tell us, for example, where to put in new roads or bridges in, what impact infrastructure construction is having on day-to-day lives and where we need to intervene to improve transportation for the population.



Source: Data-for-Good @ Real Impact Analytics (2016)

- **Distribution.** Any retailer or distributer active in B2C needs to optimize its distribution network. Mobile data can provide dynamic insights on population movement and the need to secure inventories by location on a daily basis.



Source: Global Pulse (2013) Mobile Phone Network Data for Development.

What's Next with this Module?

This is still very much a developing field and there are many questions about how to leverage mobile data. This document is a work in progress and represents an initial draft that will further develop based on user and expert feedback and as new learning occurs through further implementation. A continued framework for this document includes a second part that will detail the complexities involved in operationalizing this approach, and the precautions to bear in mind. More detailed analysis of these questions are forthcoming and will be provided at a later stage as part of this module. Topics covered will include:

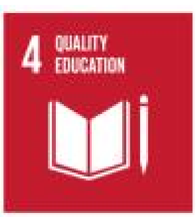
- Privacy and regulation
- Financial flows, funding mechanisms and economic relations
- Role of the NSOs
- Governance
- Legal and institutional relations
- Human and technical capabilities

Appendix 1: Examples of use cases for the SDGs

End poverty in all its forms everywhere. Emerging countries have their own poverty methodologies, used to identify their position on various regional and international rankings. Mobile data could complement this, through proxy indicators related to access to health care, services, connectivity and infrastructure, for example, disaggregated by gender and age, at high levels of granularity. Developing a multidimensional poverty model would take a considerable amount of input data, but poverty can be approximated by observing mobile phone usages and mobility at regional level and extrapolated at more granular level poverty maps showcasing multiple perspectives can provide policy-makers with better insights for effective responses for poverty eradication.[3]

End hunger, achieve food security and improved nutrition and promote sustainable agriculture. We can monitor the risk of food crisis by focusing on activities on sensitive antenna in sensitive markets. For example, lower purchase of top-ups may correlate with lower food spending in market-dependent households, and thus indicate a situation of lower food security, and at least a demand for human interpretation of the circumstances to assess the need for intervention.

Ensure healthy lives and promote well-being for all at all ages. Mobility data can be proxied to develop epidemiology models to measure and prevent disease spread. Mobility data show us who is moving and in what numbers, helping us calculate a disease vector and predict how it will spread. It can also help us monitor quarantine effectiveness. The analysis can then be triangulated for more precise intervention.



Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all. We can use mobile data to measure the percentage of youth/adults with ICT skills, by type of skill. Leading on from this, we can then proxy monitor the literacy rate. Ratios of SMS and voice calls are highly correlated with the literacy rate, and can be connected with maps and ethnicity data to produce information that can be used to design and localize interventions more efficiently.

[3] See Pokhriyal, N., Dong, W. and Govindaraju, V., National Statistics Prize: Virtual Networks and Poverty Analysis in Senegal.

Achieve gender equality and empower all women and girls. It is possible to measure in the metadata a number of key variables that can determine with a high level of accuracy the gender of the user of a phone. We can use this to measure indicators such as the proportion of individuals who own a mobile telephone, by sex. We can use these data in a number of ways, such as to directly feed banking or mobile money tools that aim to offer or facilitate access to loans for female populations.

Ensure availability and sustainable management of water and sanitation for all. On certain antennas, if there is a drought, it will be possible to see a great deal of phone activity around wells and water sources. Although there may be reasons other than drought for such activity, it will now be possible to raise a query and interpret the location and mobility data in a human way to ensure interventions are possible if necessary. It is also possible to monitor weather conditions, and then to use the emerging data to make decisions related to agriculture. For example, agricultural experts at the center will be able to pass on weather information to farmers in remote rural areas via SMS.

Ensure access to affordable, reliable, sustainable and modern energy for all. Along with measuring the percentage of the population with electricity access, we can develop a model for optimal network deployment using the correlation between mobile and energy usage. Mobile phone data are an accurate proxy of energy needs and can be used to develop bottom-up demand models that prioritize electrification in areas with scarce information on local activities and energy consumption.[4] We can also help reduce the cost and increase the quality of electricity by feeding real-time load management.

Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all. We can measure the percentage of adults with an account at a financial institution or with a mobile money service provider: using CDRs we can measure mobile money penetration and usage geographically and over time. We can also measure the number of accounts mobile money services reach—although we cannot track accounts without access to data from financial institutions.

In terms of one particular growth sector, we can observe tourism movements towards a country and within it, by tracking roaming codes and calls. These data will be useful for national statistics and for economic planning, as well as for sector-specific planning and programming. We may also be able to analyze the links between economic growth, social structure and mobility.

[4] See Martínez-Ceseña, E.A., Mancarella, P., Ndiaye, M. and Schläpfer, M. First Prize and Energy Prize: Using Mobile Phone Data for Electrification Planning.

Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation. We can analyze the need for investments in infrastructure, such as dams, electrification, roads and railways. Anonymous phone traces can be altered with an algorithm to predict future mobility. Information is used to improve decision-making for road network planning,[5] for example where to put in new roads or bridges, what impact infrastructure construction is having on day-to-day lives and where we need to intervene to improve transportation for the population.

Reduce inequality within and among countries. Mobility data can also work against exclusion, including through increased digital penetration in remote areas and for marginalized groups. We can also use air-time spending as a proxy for poverty so we can identify the bottom 40% of all mobile users and track over time the evolution of expenditure as a proxy for growth, so as to design interventions that answer to needs.

Make cities and human settlements inclusive, safe, resilient and sustainable. We can measure the population with convenient access to public transport, disaggregated by age, sex and disability, by accessing transportation system data together with people's home location. Traffic analysis and assessment of time taken to move from one key point to another, at different times and under different circumstances, can also help us improve urban planning.

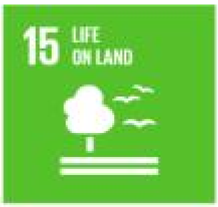
Ensure sustainable consumption and production patterns. Measure proportion of recycled phones. Additional use cases will be captured.

Take urgent action to combat climate change and its impacts. In relation to the SDG indicator on the number of deaths, missing people, injured, relocated or evacuated as a result of disasters per 100,000 people, using CDRs we can estimate the mobility patterns of people over time. This allows us to compute the number of relocated and evacuated people because of disasters (if we know when and where the disaster took place). We can also measure rain/weather crisis via radio links.

---

[5] See Y. Wang, G. Homem de Almeida Correia and E. de Romph, Transport Prize: National and Regional Road Network Optimization for Senegal Using Mobile Phone Data

Conserve and sustainably use the oceans, seas and marine resources for sustainable development. Monitor Fisherman areas economic activities. Additional use cases will be captured.



Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss. Monitor activities in non-populated areas. Monitor rain volumes through Radio links

Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels. Detect violent events, Contribute with services on ID.

Strengthen the means of implementation and revitalize the global partnership for sustainable development. On the SDG indicator on the proportion of individuals using the internet, call detail records can be used to estimate the number of subscribers accessing data per geographic region. By combining this with CRM data, we can further disaggregate the numbers by gender. With CDRs we can monitor the activity of subscribers, which allows us to account for inactive subscribers. We can also measure trends (increase, decrease) in internet usage over time.