Using Mobile Data for Development

May 2014

Cartesian™

BILL & MELINDA GATES foundation

About Cartesian

Cartesian is a provider of strategy consulting, professional services and managed solutions serving the global communications, technology, and digital media industries. We combine our deep understanding of the global market with creative, yet rigorous, analytic techniques to provide strategic management consulting services to companies around the world. Our offices are located in Boston, Kansas City, London, New York, and Washington.

About the Financial Services for the Poor Team at the Bill & Melinda Gates Foundation

A growing body of evidence suggests that increasing poor people's access to better financial tools can help accelerate the rate at which they move out of poverty and help them hold on to economic gains. The Bill & Melinda Gates Foundation's Financial Services for the Poor program aims to play a catalytic role in broadening the reach of digital payment systems, particularly in poor and rural areas, and expanding the range of services available on these platforms.

About the Authors

Ed Naef is Vice-President of Strategy Consulting in Cartesian's Boston office.
+1-617-999-1138
Email: ed.naef@cartesian.com

Philipp Muelbert is a Principal in Cartesian's Boston office.
Email: philipp.muelbert@cartesian.com

Syed Raza is a Consultant in Cartesian's Boston office.
Email: syed.raza@cartesian.com

Raquel Frederick is a Business Analyst in Cartesian's Boston office.
Email: raquel.frederick@cartesian.com

Jake Kendall leads the Research and Innovation initiative within the Financial Services for the Poor team at the Bill & Melinda Gates Foundation.
Email: jake.kendall@gatesfoundation.org

Nirant Gupta is an Innovation Associate in the Research and Innovation initiative within the Financial Services for the Poor team at the Bill & Melinda Gates Foundation.
Email: nirant.gupta@gatesfoundation.org

We would like to thank **Katherine Cali** at Cartesian for her help. We would also like to thank the following individuals for contributing their deep expertise and thoughts to this document: **John Villasenor** of UCLA and the Brookings Institution; **Tilman Ehrbeck** and **Greg Chen** of CGAP; **Chris Locke** of Caribou Digital; **Robert Kirkpatrick** of UN Global Pulse; **Andy Tatem** of University of Southampton and Flowminder.org; **Linus Bengtsson** and **Erik Wetter** of Flowminder.org**; Joshua Blumenstock** of University of Washington; **Sendhil Mullainathan** of Harvard.

Contents

1. EXECUTIVE SUMMARY

Mobile phones have become ubiquitous, used not only by relatively wealthy consumers in developed markets but also increasingly by people in the world's poorest countries. In 2012, there were 5.9 billion active mobile connections globally which has been forecasted to increase to 7.6 billion in 2017[1]. As the power of mobile devices has increased and their cost has fallen, more and more people around the world have found them to be critical tools that enhance their daily lives.

Mobile devices generate a range of data about their users. Information about identity, location, social patterns, movement, finances and even ambient environmental conditions can be derived from the data logged in mobile systems. As this data is uniquely detailed and tractable, it can capture information not easily found from other sources at a scale that would be difficult to recreate through other means. In particular, mobile is one of the only large-scale, digital data sources that touch large portions of low income populations in developing countries, implying that solutions identified in one market can easily be experimented with in another. While this data is personal and private, if it is analyzed under proper protections and anonymization protocols, it can be used to enhance the lives of poor people around the world across a range of dimensions.

For example, mobile data has been used by researchers, mobile operators and governments to help plan emergency response after natural disasters, enhance access to financial services for the poor, track the spread of infectious disease, and understand migration patterns of vulnerable populations. Indeed the full range of ways that mobile data can be used to improve the lives of poor people is only beginning to be explored.

Economic development projects require keen insight into people's lives—their habits and behaviors, their health and prosperity levels, and their needs and aspirations. It is critical to clearly understand the problems of the poor before trying to help fix them. Therefore, having granular data that captures the experiences of poor communities, along with the analytical techniques needed to decipher that data, allows researchers and development practitioners to improve the accuracy, effectiveness, and reach of their initiatives. Practitioners in the field of economic and social development can better monitor and track the progress of their programs in almost real time, bring projects to scale at a lower cost, gather rapid feedback from the field, collaborate more effectively with stakeholders, and demonstrate impactful outcomes.

While the opportunity to use mobile data for development goals is increasingly accepted, challenges and barriers remain. Data needs to be available and accessible. It needs to be presented in a format that can be understood and utilized. Operators' commercial objectives need to be considered. Most importantly, data must be used in a way that does not infringe on data protection rules and an individual's right to privacy.

Many operators, researchers, and governments have explored ways to deal with these challenges—some more successfully than others—through anonymization, aggregation, opt-in/opt-out models, regulations, and legislation. The areas of privacy and data sharing are especially critical and are evolving every day, meaning it will take time for a consensus to be found. Still, many of the relevant parties are gradually coalescing around uniform practices.

[1] Wireless Intelligence, Global Mobile Penetration- subscribers vs. connections.

The net result of these barriers is that while compelling development studies and demonstration projects have been completed with mobile data, too infrequently efforts do not scale or are not repeated. Data sets age and are not refreshed. Finally, access to data sets often needs to be negotiated with operators and regulators on a case-by-case basis or in pre-defined situations due to lack of accepted sharing best practices.

Before conducting research with mobile data, one must understand mobile networks and data stores in some detail. This paper aims to increase awareness amongst researchers and development leaders of the power of mobile data for development goals, to illustrate in more detail how data is gathered, what type of data exists in mobile networks, and to explore ways to overcome barriers to its use. Specifically this paper:

- Provides an overview of how people in the developing world adopt and use mobile phones.
- Illustrates how these behaviors generate mobile data that can be extracted from mobile networks, what kind of data is available, and how it can be interpreted.
- Identifies some of the ways mobile data can be leveraged for a range of development goals.
- Presents potential obstacles to using mobile data and ways to remove those barriers so it can be made accessible in appropriate ways.
- Provides an overview of key privacy rules and some of the ways to address the important requirement of maintaining user data privacy.

This work was conducted as a joint research project between Cartesian and the Research and Innovation Initiative of the Financial Services for the Poor team at the Bill & Melinda Gates Foundation.[2] The facts and findings in this paper are derived from more than 50 primary interviews conducted with operators, technology service providers, regulators, banks, researchers and academics, and industry associations, as well as from a review of a number of academic, industry, and policy studies and Cartesian's experience as a consultancy specialized in communications markets and technology. The research for this project was conducted primarily in the second half of 2013.

Our research focuses on 10 countries in sub-Saharan Africa and emerging Asia, but our findings on the broad availability and possibility of mobile data applications have universal applicability. The countries we selected to study include Kenya, Tanzania, Uganda, Nigeria, Pakistan, India, Bangladesh, Indonesia, Botswana, and the Philippines. These countries, which are currently the targets of a broad spectrum of development efforts, were chosen to provide a range of geographic and income profiles. On average, more than 60 percent of residents in these countries live on less than US $2 a day, which is a common metric for defining "global poverty" and the focus population for the Bill & Melinda Gates Foundation.[3]

While comprehensively covering all of these subjects in any single paper would be challenge, we seek to strike a balance between a detailed description of our findings and providing an overview for a variety of audiences. Our hope is that this paper builds on the foundations laid down by so many key stakeholders in this field and accelerates the appropriate use of mobile data to improve the lives of poor people around the world.

[2] This research is the product of the authors and does not necessarily represent the views of the Bill and Melinda Gates Foundation.

[3] Development Research Group, Poverty Headcount Ratio at $2 a Day (PPP) (% of Population), World Development Indicators (Washington, D.C.: World Bank, 2013), http://data.worldbank.org/indicator/SI.POV.2DAY.

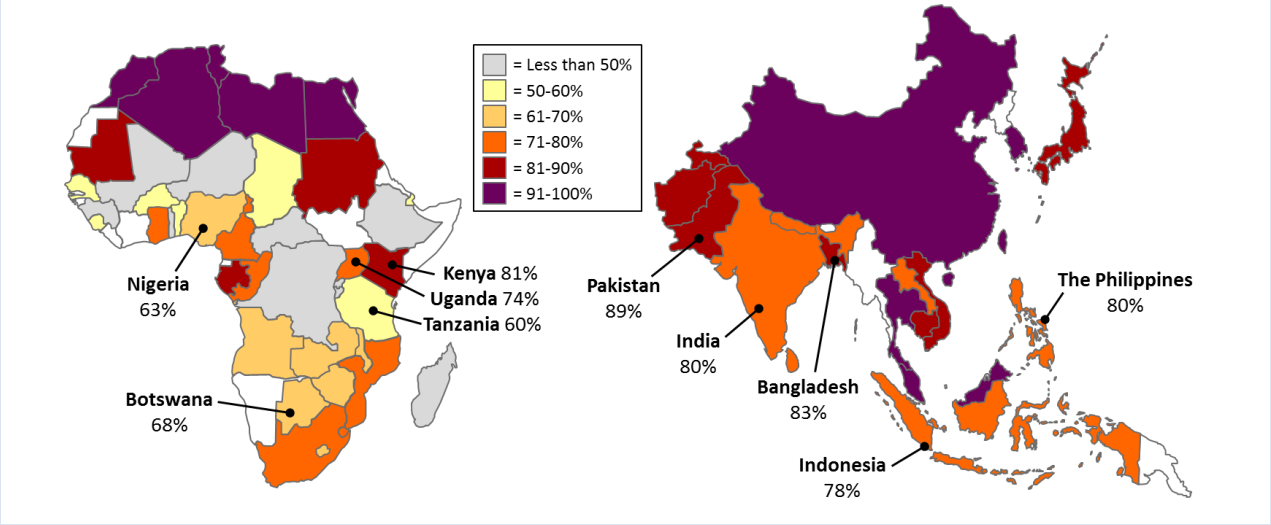
2. Adoption and Usage of Mobile Phones in Developing Countries

Mobile phone usage has increased in recent years, with mobile phones now in the hands of or accessible to more than 70 percent of the population in most low- and middle-income countries. While mobile phone availability is lower among poor and vulnerable populations, it is not substantially lower.[4] The total cost of ownership of mobile phone usage is falling and poor people increasingly view mobile phones as a necessity rather than a luxury. While most low income populations are primarily using basic mobile feature phones for text messaging and voice calls, faster mobile networks and falling devices costs are catalyzing mobile internet and advanced "smart phone" penetration.

2.1. Mobile Adoption Rates Are High in Developing Countries

In recent years, rates of mobile adoption have grown rapidly worldwide, with the fastest growth occurring in developing countries. There were 6.8 billion active global subscriptions in 2013, nearly 96 for every 100 people on the planet. Mobile phones are similarly ubiquitous in the developing world at 89 active subscriptions per 100 people, though because people can own multiple subscriptions and share subscriptions across multiple people, these numbers are only roughly indicative of the rate of mobile ownership in the population.[5] However, household surveys confirm high rates of ownership and phone usage in the developing world as the 2012 Gallup World Poll suggests (see Figure 1 for examples from Africa and Asia).

Figure 1 – 2012 Gallup Poll, "Does your home have a cellular phone?" (Ages 15 and over)



Source: Gallup. Note: Data not available for countries in white. 2011 data used for Angola, Botswana, Burundi, Central African Republic, Chad, Democratic Republic of the Congo, Djibouti, Gabon, Guinea, Lesotho, Madagascar, Malawi, Mali, Mozambique, Niger, Republic of the Congo, Rwanda, Sierra Leone, Swaziland, Togo

In African and South Asian countries, including the countries of our study, actual rates of phone ownership are usually greater than 50 percent among adults (14 years and older), and a further 10

[4] Source: Kendall et al, *Remittances, Payments, and Money Transfers: Behaviors of South Asians and Indonesians and Payments and Money Transfer Behavior of Sub-Saharan African Households;* Gallup whitepapers.

[5] Source: ITU; *The World in 2013: ICT Facts and Figures.*

percent to 20 percent report using the phone of their neighbor, friend, or family member.[6] (See Figure 2.) While only half of adults currently own mobile phones, access through family and social relationships and continued penetration of mobile devices will increase mobile reach in the short term.

Mobile SIM Penetration Figures vs. Household Survey Data

Because of the focus on subscriptions rather than numbers of people reached, the most common indicator used in industry circles is the mobile penetration rate, defined as the ratio of the number of mobile service subscriptions (SIM cards) in a country to its total population. However, this rate does not accurately reflect the proportion of adults who regularly use mobile service because in most developing countries, many people tend to have more than one active SIM subscription and some share subscriptions. Many savvy callers in the developing world regularly swap out their SIM cards in order to maximize network coverage, to take advantage of the latest offers by carriers, or to avoid the extra fees of inter-carrier calls. Upwards of 95 percent of mobile connections in these markets are prepaid, further complicating accurate measurement of mobile ownership and adoption.[7] Definitions of what constitutes an "active" SIM for pre-paid customers vary from operator to operator, leading to further inflation in penetration rates.

- In Indonesia, high interconnect rates and on-net/off-net tariff differentials[8] have resulted in an average of 2.62 SIMs per user.[9]
- In India, as much as a quarter of total connections of almost 1 billion were considered inactive by the regulator in Q2 2012.[10]
- In Africa, it is estimated that there are only half as many unique mobile subscribers as there are mobile subscriptions, meaning each mobile subscriber has an average of two SIM cards.[11]

In contrast, household surveys conducted by market research firms or economics researchers ask nationally representative samples of users directly about their mobile phone ownership and usage. While household surveys have their own challenges, including over reporting of phone ownership by people who may be embarrassed to admit they don't own one, they are still the best source of data to infer the percent of population who has access to mobile phones.

[6] Gallup. Classification: No access to mobile phone, use mobile phone of friend, neighbor or family member, own a mobile phone. Although different sources indicate different mobile ownership / access levels, all confirm a substantial and meaningful level of mobile adoption for development purposes.

[7] GSMA Mobile Intelligence.

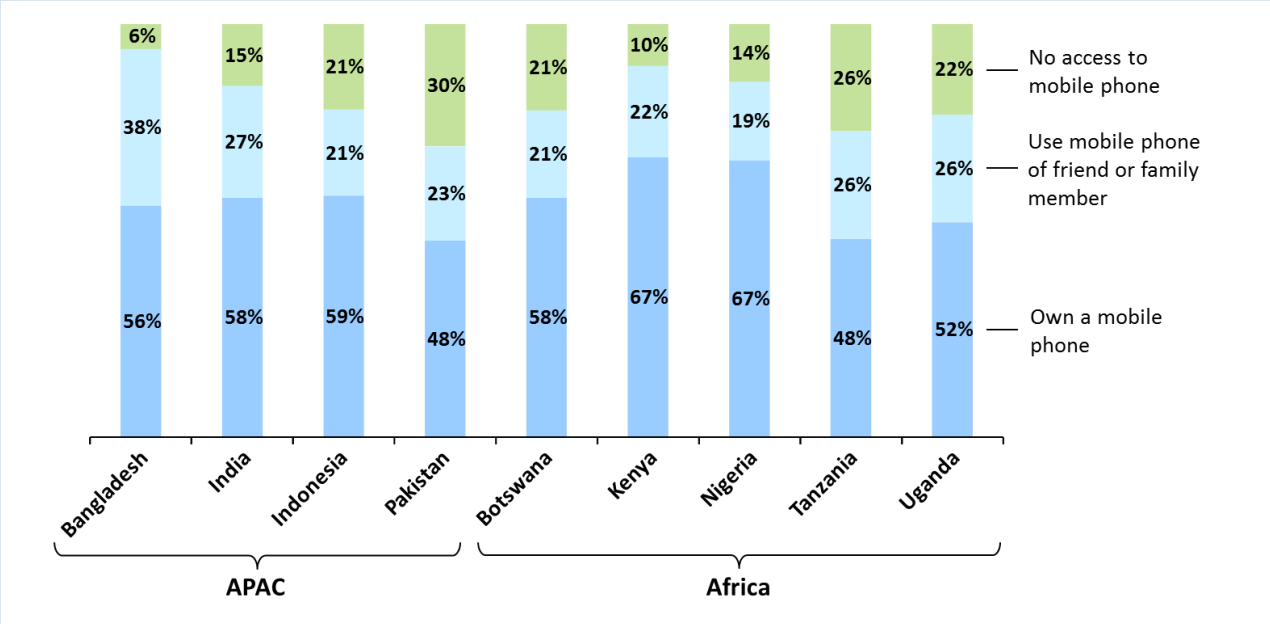
[8] Credit Suisse.

[9] Wireless Intelligence, 'Global mobile penetration — subscribers versus connections', Joss Gillet, October 2012.

[10] Wireless Intelligence, 'Global mobile penetration — subscribers versus connections', Joss Gillet, October 2012.

[11] Wireless Intelligence, 'Global mobile penetration — subscribers versus connections', Joss Gillet, October 2012.

Figure 2 – Mobile Phone Ownership and Usage Among Adults Across Target Countries in 2012



Source: Gallup

Note: Complete data for The Philippines not available. Personal phone ownership is 50%

It should be noted that in developing markets, mobile adoption by women is 20 percent or more lower than that of men.[12] The typical size of the gap, as with other measures of gender inequality, varies significantly by country and region. When women do have access to mobile service, they are more likely than men to rely on a handset they do not own personally, such as a household phone or one belonging to a husband or other family member. The source of this gender gap lies in both financial and cultural constraints. However, women who do own, or want to own, mobile phones emphasize their empowering impact: mobile phones increase women's sense of safety, social connectedness, independence, and economic opportunity.[13]

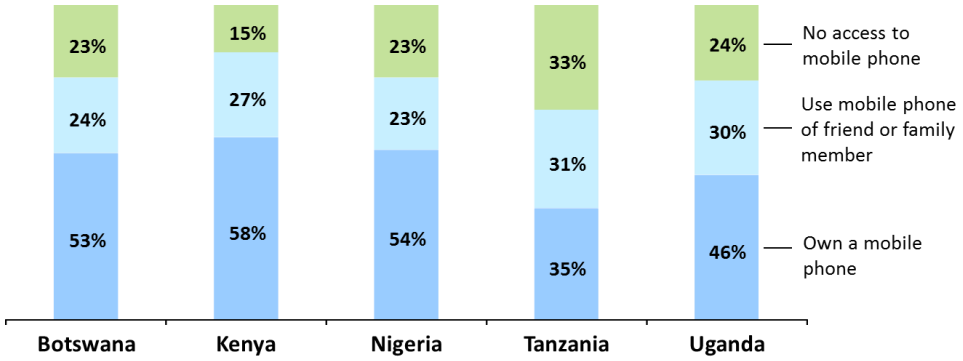
2.2. Rates of Mobile Ownership and Usage Are Equally High among the Poor

Mobile adoption and usage rates are surprisingly high even among the poorest residents of developing countries. In our 10 target countries, the majority of people living in poverty either have their own mobile phone or have some degree of regular access to mobile service through other people's devices. Some own only SIM cards and share devices with friends or relatives. (See Figures 3 and 4.)

[12] GSMA/ Cherie Blair Foundation.

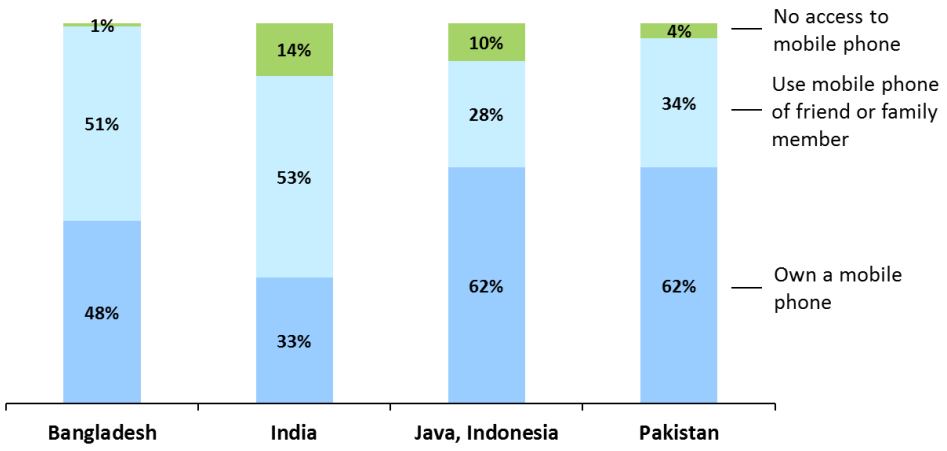
[13] Samarajiva R, "How the Poor Use ICTs: Findings From Multi-Country Studies of Teleuse at the Bottom of the Pyramid," presented at the Regional FAO Workshop on the Use of Mobile Technologies in Agriculture, Bangkok, Thailand, April 3, 2012, http://lirneasia.net/wp-content/uploads/2012/04/Samarajiva_FAO_BKK.pdf.
Vital Wave Consulting, *Women & Mobile: A Global Opportunity: A Study on the Mobile Phone Gender Gap in Low- and Middle-Income Countries,* London, UK: GSMA Association & Cherie Blair Foundation for Women, 2010. Crandall et al., 'Mobile Phone Usage at the Kenyan Base of the Pyramid.'

Figure 3 – Mobile Access Among People Making $1 a Day or Less



Source: Gallup, 'Payments and Money Transfer Behavior of Sub-Saharan African Households', 2012, BMGF.
Note: Gallup classification: No access to mobile phone, use mobile phone of friend, neighbor or family member, own a mobile phone.

Figure 4 – Mobile Access Among People at the Bottom of the Pyramid (SEC Groups D and E)

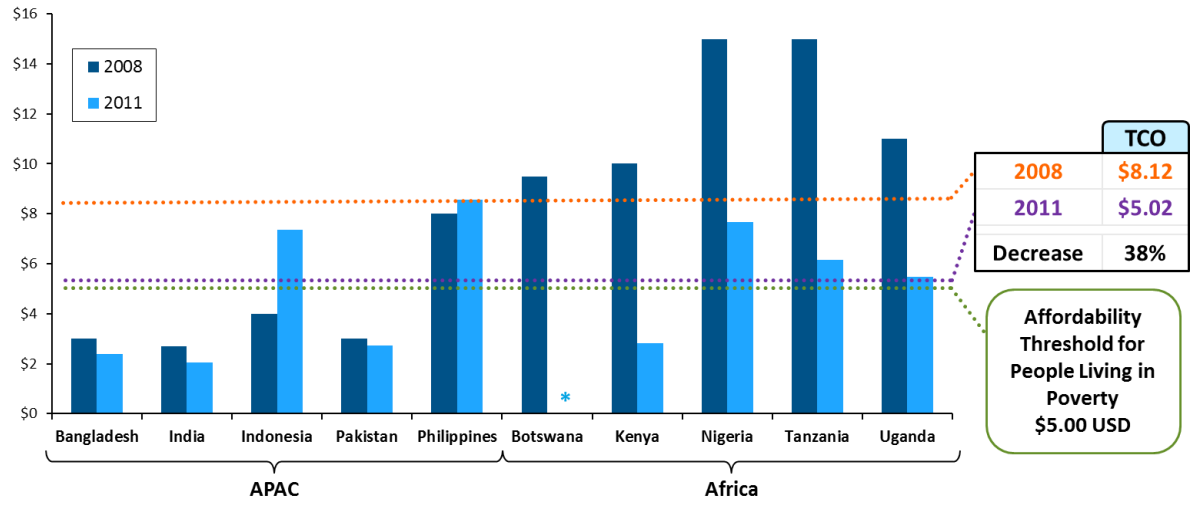


Source: Samarajiva/LIRNEasia, "How the poor use ICTs."
Note: BOP defined as Socio-Economic Classification (SEC) Groups D and E. Indonesia is partial data for country.
(http://lirneasia.net/wp-content/uploads/2012/04/Samarajiva_FAO_BKK.pdf)

This high level of adoption can be attributed to two primary factors: affordability and necessity. The average total cost of mobile ownership (TCO) in our selected countries has dropped by almost 40 percent from 2008-2011, according to an estimate by Nokia Research. A TCO of US $5.00 per month could make mobile phones affordable to even the poorest.[14] (See Figure 5.)

[14] GSMA Mobile Intelligence; Nokia/Nokia Siemens Network, "Knocking Down the Affordability Barrier," Expanding Horizons, February 2009.

Figure 5 – Monthly Total Cost of Mobile Ownership, 2008 and 2011



Source: Nokia Research, GSMA Mobile Intelligence.
Note: Indonesia and Philippines showed a rise due to increased interconnect rates.
* Information not available for Botswana in 2011.

In terms of need, many poor people in developing countries see mobile service as a necessity rather than a luxury. Among the poorest, spending on mobile service varies little with income, showing an inelasticity that is characteristic of essential goods. For example, in one study conducted across several APAC countries people spent about the same amount on mobile service whether it cost them an average of 2.5 to four days of one household member's income or more than a week's.[15] Mean monthly mobile spending only varies minimally across low and lower income groups, and sometimes displaces spending on other "essential" items for some of the lowest income groups.[16]

Perhaps to an even greater extent than better-off phone owners, the poor value mobile service for its social and financial benefits. For example, one in five Kenyans living in poverty reported forgoing or reducing a usual expense, almost always food, to pay for phone credit. Eighteen percent of poor Kenyan mobile subscribers reported that they had earned more money because their phones made them more "reachable" to employers. In five emerging APAC countries, poor people most often pointed to reduced travel time, improved relationships, and, above all, increased ability to respond to an emergency as the top benefits of phone access. Many poor people in developing countries now rely on mobile service as an essential tool for managing their personal and working lives.[17]

Still, for people living in poverty throughout the world, service and handset costs remain a major barrier to wider adoption.

[15] Agüero A, de Silva H and Kang J. "Bottom of the Pyramid Expenditure Patterns on Mobile Services in Selected Emerging Asian Countries," Information Technologies & International Development 7(3): 19–32, 2011.
[16] Agüero, et al; Crandall et al., 'Mobile Phone Usage'.
[17] Crandall et al., *Mobile Phone Usage;* Samarajiva, "How the Poor Use ICTs."

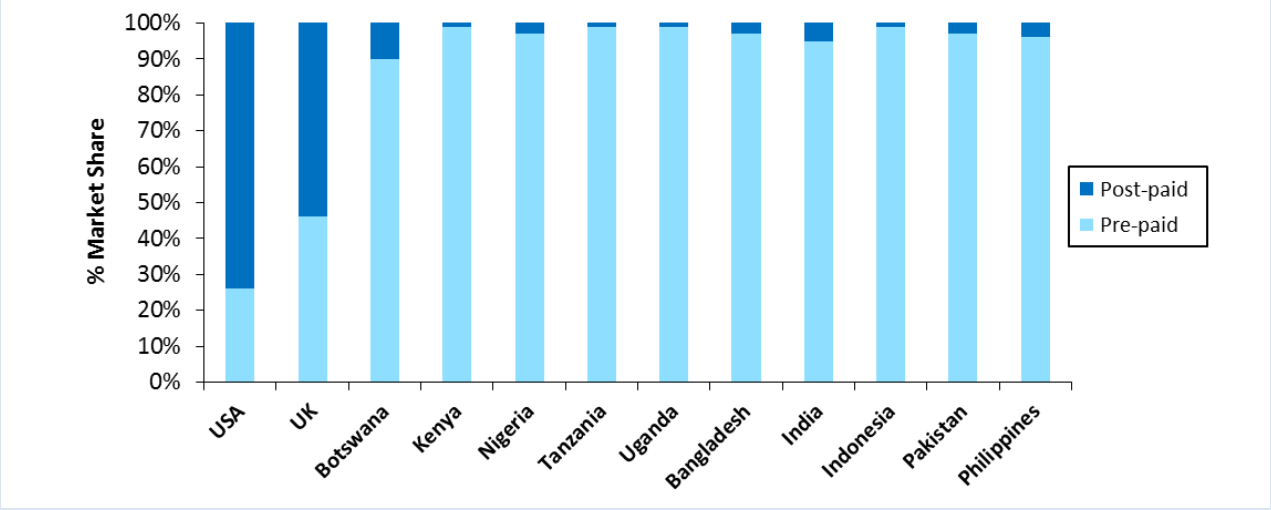
2.3. Usage Patterns in Developing Countries Differ from Those in the Developed World

While adoption and usage rates are often as high in the developing world as they are in the developed world, usage patterns are different, which means that the data captured by mobile operators about subscribers is different too. For example, people in developing countries use more voice and SMS text as compared to data, and this affects the nature of information that is captured by operators.

As mentioned above, many people in developing countries share phones and/or SIM cards, sometimes regularly and sometimes in cases of emergency. Because of this, data analysis efforts that rely on individual-level mobile data (as compared to household-level, for example) will need to take these usage patterns into account in the developing world. Multi-SIM ownership, as discussed above, also affects the nature and granularity of mobile data available from developing world subscribers.

Another key difference is that 2nd generation (2G) networks are the dominant technology in many developing countries and will be over the near to medium term.[18] Built mainly for voice and text services, 2G networks enable calling, SMS and services such as informational inquiries, purchasing ringtones, listening to music, and access to mobile money. People living in poverty typically buy pre-paid plans (Figure 7) and use basic or feature phones (that today have media players/ FM radios, cameras, basic Internet access and sometimes GPS) that run on 2G networks and cost $10 to $25.

Figure 7 – Percentage of Pre-Paid vs. Post-Paid Mobile Subscriptions



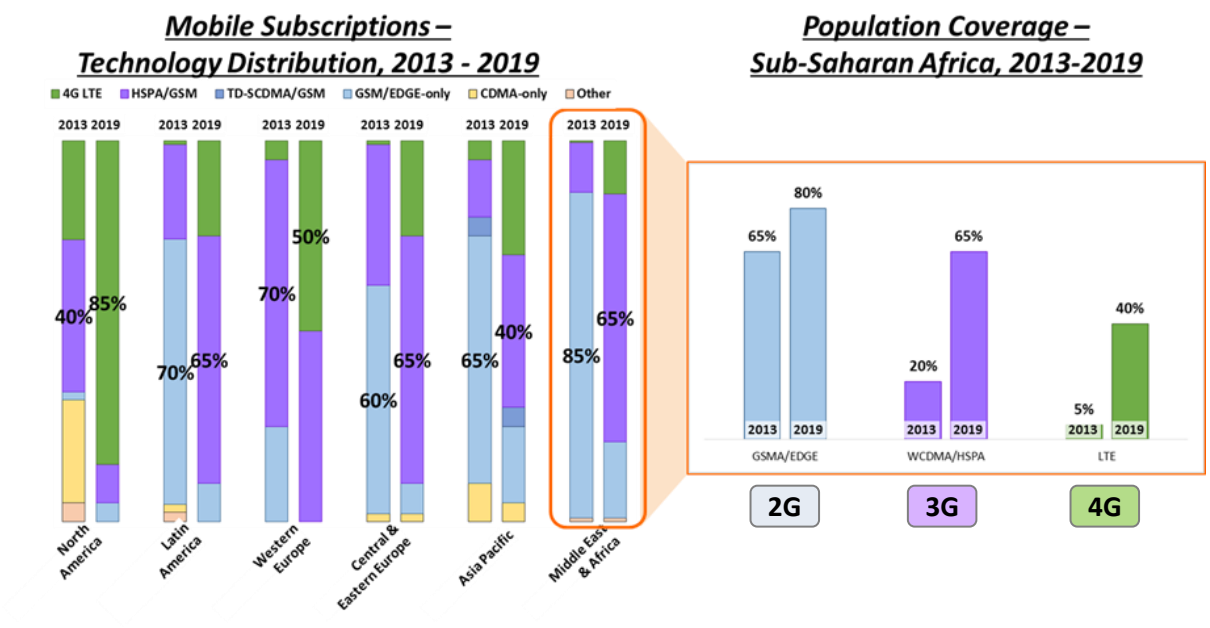
Source: GSMA Mobile Intelligence, Agüero, de Silva, and Kang.

That being said, in the majority of our target countries operators are investing heavily in upgrading 2G networks to faster 3G+ networks, although the speed of transition will vary across markets over the next three to five years. By 2019, more than half of the Sub-Saharan population will be covered by 3G services while 2G networks will cover ~80% of the region's inhabitants. (Figure 8). 3G networks provide a wider range of voice and broadband data services, while 4G networks further expand the possibilities of mobile applications by delivering communications four to 10 times faster than 3G networks. The rapid migration to 3G+ networks in many countries makes the collection of richer datasets possible and allows

[18] Ovum, 2012-2017 Forecast.

for smarter devices to be used. Distinguished by their advanced computing capacity and connectivity that allows them to run full operating systems, smartphones allow users to take full advantage of these speedier networks.

Figure 8 – Network Connections and Coverage by Network Generation



Source: Ericsson Mobility Report, June 2014

Note: Population coverage is the proportion of the population in an area that has sufficient signal to connect to a mobile network. Ability to utilize the technology is subject to other factors as well, such as access to devices and subscriptions.

Smartphone adoption is significant, (Figure 9), and will accelerate as handset prices continue to drop, with major manufacturers targeting a $25 price point in two years, from the current $80-100.[19] The expansion of next generation networks and smartphone adoption will create additional opportunities for data capture and analysis in the future, as well as a platform for mobile operators to expand their value-added service offerings. That said, it will take a number of years for smart phones to reach a majority of the low income population.

[19] Samsung, Huawei, GSMA, Nextweb. Mozilla debuted a smartphone they claimed would retail for $25 at the 2014 GSMA mobile world congress.

Figure 9 – Smartphone Penetration by Country (2013)

Country	2013 Smartphone Subs (MMs)	2013 Smartphone Subs Growth	2013 Population Penetration
China	422	26%	31%
India	117	55%	10%
Brazil	72	38%	36%
Indonesia	48	42%	19%
Mexico	22	49%	19%
Egypt	21	41%	25%
Philippines	20	43%	19%
Nigeria	20	43%	12%
South Africa	20	32%	41%
Thailand	18	27%	27%
World	1,786	28%	25%

Source: Informa, Mary Meeker

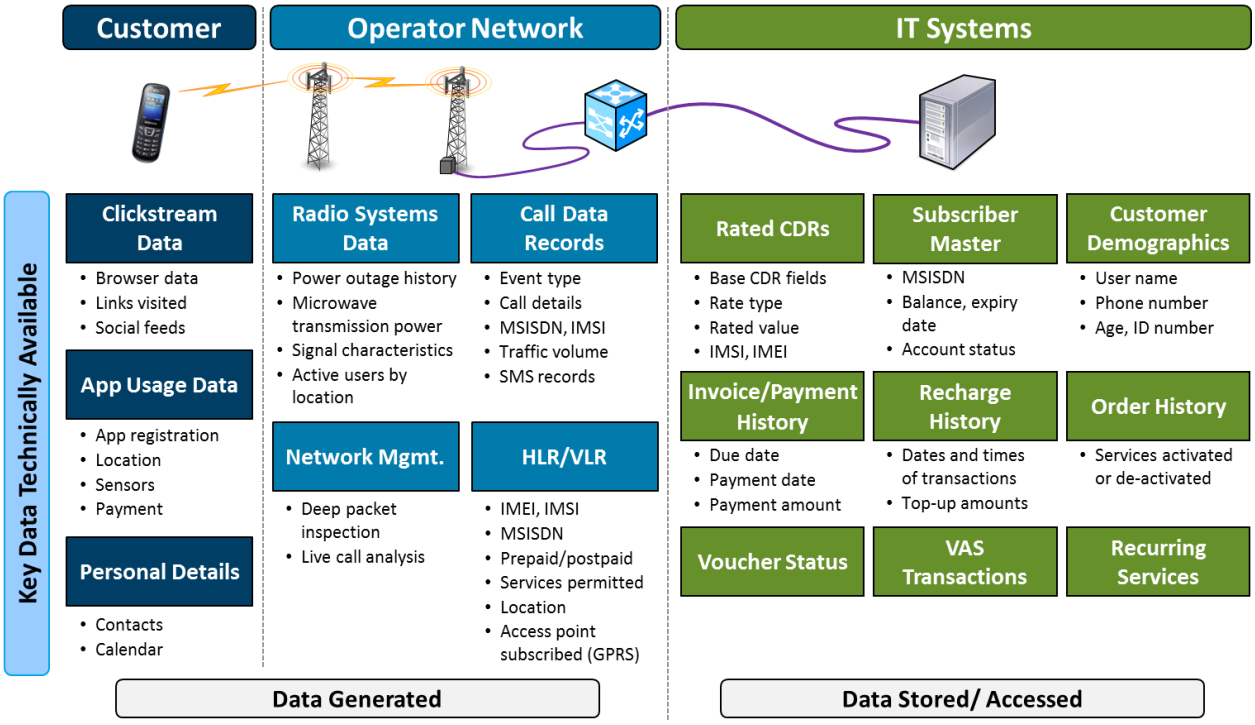
3. What Is Captured in Mobile Data Systems?

Due to the widespread adoption of mobile service in many developing countries, an extensive amount of mobile usage data is being created every day that can potentially power development initiatives. Important information about users is captured by telecommunications systems each time an operator enrolls a customer, as well as each time a call is made, a text message is sent, or a mobile money transaction is conducted. Analysis of mobile datasets can enable researchers, development professionals, and policymakers to better understand the financial, health, agricultural, educational, and other needs of the populations they serve with a level of granularity and precision difficult to achieve using other, more traditional, data sources, thus leading to improved products and services that better meet the needs of the poor.

3.1. How a Mobile Network Functions

Mobile network systems capture data from multiple sources, including the user's device; core network elements such as the cell site, switching equipment, and backhaul; and the operator's billing, mediation, customer relationship management (CRM), and data warehousing (DWH) systems (Figure 10). In order to understand the significance of each system in generating, capturing, and storing data, we must first understand how they work together to enable mobile service.

Figure 10 – The Mobile Network, Associated Systems, and Key Datasets



Customer: The customer's mobile device provides the first set of data: personal information, contacts, social graph, basic clickstream data, and user location and movement patterns. This type of information can be gleaned during active and passive usage of mobile services. As people's mobile devices become

"smarter," researchers can obtain richer insights—such as accurate location and information on diagnostic and ambient conditions—with the degree of accuracy dependent on network and device generation. For the most part, any data obtained directly from the user's device would require users to "opt in" and grant consent to access the data.

Operator Network: Mobile phone networks transmit signals carrying voice, text, and digital data from one mobile device to another through a network made of 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) is a critical element of the BSS that 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. However, this information is only available to the operators for internal network planning or to law enforcement due to data privacy considerations, as well as extraction cost considerations. We will cover these issues in Section 5, Data Protection and Privacy Considerations.

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. Figure 11 explains how attenuation from radio signals was used to collect a large amount of accurate and timely rainfall data.

Figure 11 – Estimating Rainfall Using Mobile Network Data

- Researchers from the Royal Netherlands Meteorological Institute have devised a means to use the attenuation that results from radio signals when rain falls between cellular towers to measure the amount of rain that falls in an area.
- Data from ~2,400 sites operated by T-Mobile were collected over 12 days in 2011 and provided more widespread data than a typical network of rain gauges.
- Such accurate and timely surface precipitation measurements can bring critical information at low cost to water resource managers, farmers, and climate researchers, particularly in the developing world where standard rain gauge networks are usually underdeveloped.
- In the future, cell phone towers could be used to obtain long-term (e.g., annual) rainfall maps of a region or country, which could further improve the accuracy of weather radar data.

Source: Proceedings for the National Academy of Sciences.

The mobile switching center (MSC) is the primary service delivery node, responsible for routing voice calls and SMS as well as other services. Each switching center hosts a visitor location register (VLR), which is a database containing information such as temporary IDs of the subscribers roaming within the MSC's location area[20]. Additionally, the home location register (HLR) is a central database that holds data regarding each of the cellular network's subscribers and also serves as the main source of recent location information when a call comes in and the system need to locate the user to connect the call. Whenever a telecommunications transaction (such as a phone call) occurs, the MSC generates a call detail record (CDR). A CDR is designed for billing and charging purposes, but it contains many interesting fields that can give us insights into the user, as will be discussed later. Data fields including IMEI[21] (handset identifier) and IMSI (international mobile subscriber identity or SIM card identifier) are also available at the switching center.

While the radio and switching elements of the operator network can provide critical data, operators have expressed reservations about sharing data due to technical, commercial, and regulatory reasons. We will discuss these in more detail throughout the paper.

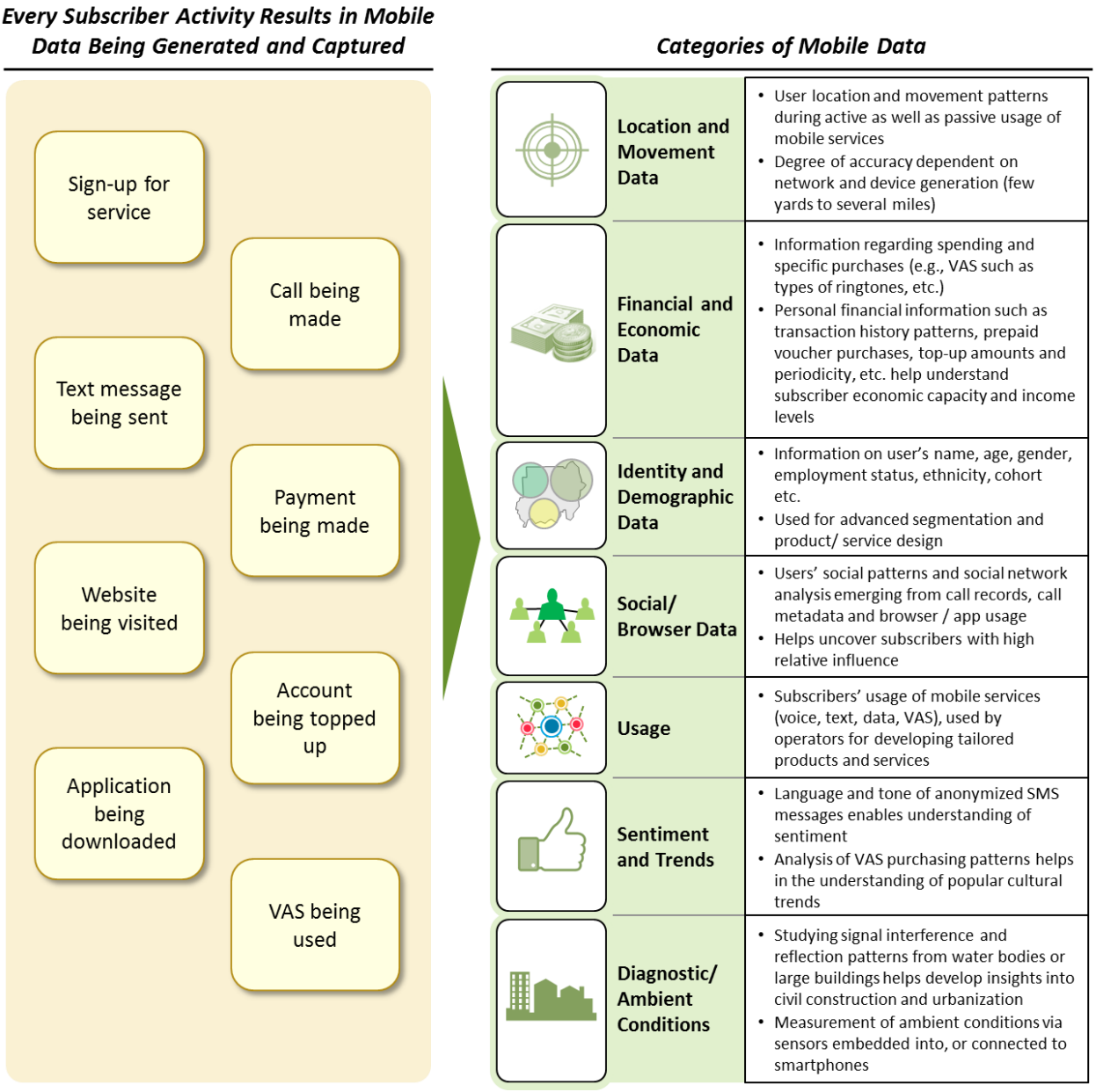
IT Systems: The cellular network also communicates with the operator's IT and back office systems, such as the systems for mediation, billing, customer relations, and business intelligence. It is important to note that the structure of operator systems tends to be more customized and nonstandard as one moves downstream from core radio and switching elements to the IT systems. For example, radio and switching elements could follow GSM (Global System for Mobile communications) specifications and are implemented across operators globally, but data warehousing, or DWH, systems are not standard. The DWH is designed to leverage the volume of customer, CDR, content, network usage, campaign, channel, product, and tariff data from multiple technologies and source systems to improve visibility for operators. DWHs are customized heavily based on the operator's business needs. Still, they serve as the best potential source of sustained data sharing with third parties due to the richness of the data and technical, business, and privacy considerations.

3.2. Mobile Data Captures a Wide Range of Customer Behaviors

Each time a user interacts with a mobile operator, many details of the interaction are captured, creating a rich dataset on consumer behaviors. Figure 12 illustrates the types of interactions with mobile systems and the types of insights they can generate.

[20] Comprised of multiple cells and managed by a single base station controller
[21] International Mobile Station Equipment Identity

Figure 12 – Customer Transactions Generate Mobile Data



There is a wide variety of information that can be learned about users from their mobile data, but three of the most important types of information include:

- **Location and mobility**: Location is tracked passively when users' phones interact with towers in each cell cite they visit, and actively, each time a user initiates a voice call, SMS, or other transaction.
- **The social network**: Calling and SMS patterns create a lens into a person's social network, including who they communicate with, how long, and how often. Further, in many emerging markets the originator of a phone call pays for the minutes of the call; even understanding who

someone calls vs. who calls them can give a sense of social stature among a social network, important for marketers seeking to reach nodal hubs of influencers

- **Recharge and purchase history**: Patterns of recharging minutes and purchases of VAS (value-added services) can give insights into an individual's economic circumstances and the financial shocks or difficulties they face.

While there are several high-value data fields that provide relevant information, including IMEI, IMSI, and usage rates of VAS, some of the most useful data elements come from CDRs. A CDR is a call detail record that is generated for every user event on the network and contains a broad spectrum of relevant information, including the phone numbers of both the calling and receiving parties, the tower location, the start time of a call, and its duration (Figure 13). CDRs are a subtype of event detail records (EDRs), which contain information about calls, SMS, value-added services, data usage, and all other services subscribers use. CDR as a term is often used to describe a phone usage data record beyond just placing and receiving calls and is used in that way in the remainder of this paper.

Figure 13 – Example Set of CDR Fields

Calling Party Number	Called Party Number	Caller Cell ID	Location Area Code	Call Time	Call Duration	Served IMSI
25358XXX	25358XXX	5102	41708	2012-12-06 14:24:32	01:25:34	4600013511068690

Note: This is a coarse representation of a complete CDR. Many typical fields are stripped for brevity.

In the near term, the CDR will remain the single most important source of information for research efforts. To date, most academic and development use cases have relied on CDRs because they offer a number of advantages. Since they follow fairly standard formats across operators, CDRs are relatively simple to aggregate, analyze, and interpret, while outputs of analysis are potentially replicable in other markets and contexts. Thanks to the range of information they include, CDRs can serve as inputs into multiple categories of mobile indicator analyses, including location, financial information, demographics, identity, social networks and sentiments, and ambient conditions analysis (Figure 14).

Figure 14 – The Many Uses of CDRs

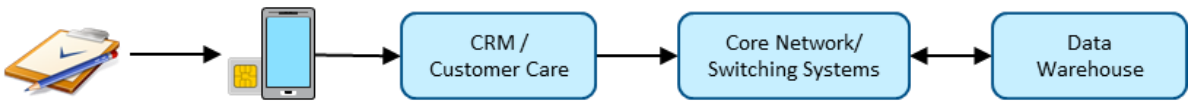
Data Categories	Key Insights from CDRs
Location and Movement Data	• Fields on Location and Cell IDs, that can be used to track population location and movements
Financial and Economic Data	• CDRs from the Mediation and Billing systems also contain spending, tariff and rating info on subscribers
Identity and Demographic Data	• Identifiers that enable monitoring of subscribers across operators and 3rd party data stores
Social/Browser Data	• Information on calling and called parties can be aggregated to develop holistic social graphs of populations to understand communication flows
Subscriber Usage Information	• Key identifiers such as IMSI, IMEI, MSISDN etc. of parties • Time of calls, duration, services activated for a given subscriber
Sentiment and Trends	• Analysis of call patterns, SMS texts or web usage history by studying CDR data

3.3. How Mobile Data Is Captured From User Interactions

Each time a user interacts with the network, both actively and passively, they "light up" the network at different points and push different types of data into operator systems. For example, an SMS text can be seen at the handset, base station, switching center, mediation, rating, billing, and the data warehouse. Below are several illustrations of the complex interplay of network and back office systems triggered by the user when they initiate a mobile service activity.

Device Activation: Know Your Customer (KYC) laws in many countries require operators to capture and store each subscriber's biographical information in an active customer database alongside information identifying the handset and SIM card the subscriber uses. Figure 15 illustrates what happens when a customer activates their device.

Figure 15 – Device Activation



Note: This simple description of the device activation scenario is intended to provide the reader with a high-level understanding of mobile network operations. In actual implementations, account activation is a complex task with multiple databases and network elements being coordinated for user provisioning.

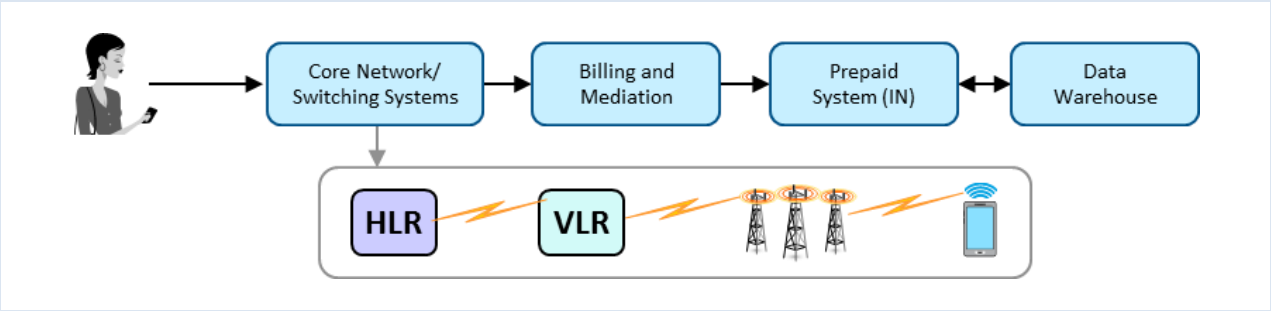
1. When the device is activated, the IMEI (handset identifier) is stored in the CRM database.
2. In the core network, the HLR (home location register) is updated to note which services the subscriber can access.
3. Also captured is the IMSI, a unique SIM identification that is used to identify details of the device in both the HLR and the visitor location register (VLR) for location and other services. When a value-added service such as mobile money is being used, the IMSI is also captured on that platform.
4. All customer profile data is processed (and archived) in the data warehouse.
5. Identifying fields such as IMSI and IMEI are also available from switch CDRs and converged billing systems.

Therefore, each time a new prepaid mobile customer registers their SIM and activates their device, the following data can be collected:

- Biographical information (e.g., name, address, age, etc.)
- Handset identifier (IMEI)
- SIM card identifier (IMSI)
- Services activated
- Allowed services
- Voucher balance

Making a Call: Figure 16 illustrates what happens when a customer makes a call.

Figure 16 – Making a Call



Note: This simple description of the call activation scenario is intended to provide the reader with a high-level understanding of mobile network operations.

1. The mobile switching center determines phone location through the HLR database:
    a. When a user initiates a call, the HLR at the Mobile Switching Center translates the dialed phone number to its unique IMSI, and then determines which VLR area the phone is located in based on the last known location.
    b. Through the towers it manages, the VLR broadcasts messages with the phone ID.
    c. The phone contacts the nearest tower and the call is initiated.
2. For postpaid customers, the billing and mediation systems are working in the background for charging purposes:
    a. After the call is complete, the mediation system receives the CDR from the core network systems and processes it for the billing systems, which are updated to reflect charges on the customer account.
3. For prepaid customers, who are the majority of people in our target countries, the process is different:
    a. The prepaid system, which is a continuous interrogation system, checks the rated call cost against the user's airtime balance that is stored in the master subscriber database.
    b. The master subscriber database is dynamic, so is continually updated with the subscriber's new balance during and after the call.

    Note: Operators are rapidly transitioning to converged billing and mediation systems that integrate prepaid and postpaid charging and other functions.
4. The CDR data is processed and archived in the data warehouse.

    Note: The retention period of mobile data, which is covered in Section 5, Data Privacy and Protection Considerations, affects researchers' ability to utilize it. For example, SMS texts could be stored at the switching element for one to seven days, moved to a data warehouse, and then transferred to tape backup where they are typically stored for one to seven years.

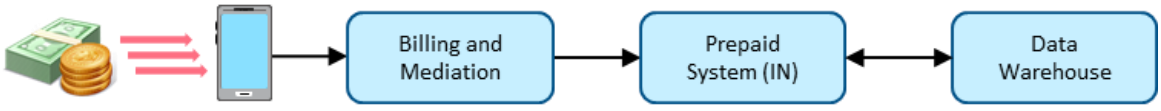
Each time a customer makes a call, the following data can be collected:

- Location
- Services activated
- Cell or tower ID, coordinates of location

- Service area code
- Handset and SIM card identifiers of both the calling and served parties
- Airtime balance
- Call type and call rate

Recharging and Topping Up a Balance: Figure 17 illustrates what happens when a customer recharges their phone and tops up their balance with prepaid vouchers.

Figure 17 – Recharging and Topping Up a Balance



Note: This simple description of the top-up scenario is intended to provide the reader with a high-level understanding of mobile network operations. Real-life implementations are understandably complex; for e.g., our research highlights that operators are moving towards converged billing systems from legacy systems.

1. Operator prepaid systems comprise many standardized databases for top-ups, resale, refills, balances, customer information, etc. A customer can recharge by using vouchers. Voucher balance, status, and dates are saved through prepaid systems:
    a. The voucher server checks the voucher number against the voucher status dataset for the value and status of the prepaid voucher entered.
    b. The new prepaid voucher is logged in the user's recharge history database.
    c. The subscriber master database is updated to reflect the user's new airtime balance.
2. Historic recharge and balance data is processed and archived in the data warehouse, where it is deemed most accessible by external parties.
3. Customer recharges can happen in many ways. For example, users can buy airtime for another subscriber through a mobile money platform. In this scenario, standard APIs (application programming interfaces) and middleware are used to communicate between the mobile money platform and the operator's top-up and billing systems.

Each time a customer recharges or tops up a balance, the following data can be collected:

- Voucher balance
- Personal identification information such as name or national ID
- Recharge dates and amounts
- Voucher number, value, status (active and expired)
- Airtime balance
- SIM card identifier

3.4. How Mobile Data Can Determine a Person's Location

Mobile data has made it possible to track the location and movements of populations. If used correctly, this is a very powerful tool that can be useful in development work in many ways, ranging from

predicting the spread patterns of epidemics to distributing resources to refugees post-disaster. In mobile operator networks, the cell ID is the simplest identifier used to determine the location of a handset. It requires the network to identify the cell tower which the device is communicating with at the time. Since the mobile handset can be anywhere in the cell area, and the cell area can be larger than 20km in diameter in rural areas,[22] the accuracy of this method can be poor. The accuracy, however, is higher in urban areas which have a denser network of smaller cell towers than in rural areas that have fewer base stations. Still, knowing how many mobile devices are in range of a tower at any point in time, and how that number changes, are powerful tools for development (e.g., to estimate local population movements).

One way operators increase the subscriber capacity[23] of a cellular network is by replacing the omnidirectional antennas at each base station with several sector antennas. Each sector is considered a new cell, thereby allowing for denser frequency reuse. This type of sectorization allows for improved location and movement tracking because of the smaller cell areas.

Location and mobility data can be ascertained in two ways: actively and passively.

1. **Passive location tracking** happens because of the inherent design of telecommunications systems. A mobile device is constantly monitoring a radio frequency channel for identification with a cell tower. As soon as a phone is turned on, a handset monitors its radio frequency channel to identify the closest cell site with the best signal in order to be prepared should the subscriber make a call. From that point on, the network keeps track of the device as it moves, even when inactive. This information is represented by the device's IMSI (SIM card identity) and logged at the VLR (visitor location register). As soon as the user moves from one location area (which is comprised of multiple cells and managed by a single base station controller) to another, the hand-off is managed by the mobile switching center (MSC) to the new MSC's VLR. Typically, VLR location information is lost as soon as the user moves to a new location area. Because of this, passive location tracking is difficult to use for analysis.

   While this type of location tracking is not always accurate because of large cell areas, accuracy can be improved by utilizing triangulation. This happens when the signal strength to the three closest towers is calculated, thereby providing a more precise location.

2. **Active location tracking** can be done in many ways. In its simplest form, when a user initiates an event such as a call or an SMS, the network captures the recipient cell ID, which is a unique number identifying each tower in an operator's network. A phone's approximate location can be determined based on the tower it is connected to. As a user travels, the change in location can be determined based on a sequence of connections made to towers along the route. Typically, such data is maintained in the CDRs which are stored anywhere from six months to seven years, depending on operator policies and government regulations.[24] In addition, smartphones allow for location tracking in numerous other ways, including global positioning system (GPS),[25] Wi-Fi access, inertia sensors, terrestrial transmitters, and Bluetooth beacons.

[22] Source: AT&T.
[23] Alcatel Lucent.
[24] Covered in more detail in Section 5.2.1 "Data Capture and Retention Requirements."
[25] GPS is now commonly available in many feature phones and can provide more accurate location information than cell-site based tracking.

3.5. Gauging the Availability and Accessibility of Data

Researchers and development practitioners need to consider both the availability and accessibility of mobile data, including which types of information can be shared, when such information can be accessed, and how it can be accessed.

In order to measure the ease of accessing certain data, the following should be taken into account:

1. Effort of extracting data
2. Criticality vs. accessibility of the system to the operator's core functions where data is stored
3. Duplication or presence in multiple operator datasets
4. Privacy, regulatory, competitiveness, and other constraints
5. Standard vs. customized representation of datasets
6. Operator policies for sharing data
7. Length of operators' data retention periods

Key identifiers such as IMSI, CDRs, recharge history, and biographical information are captured and retained at different datastores in the network and shared with third parties via multiple methods, including secure FTP channels:

- IMSI is available from switch CDRs and convergent billing systems and remains valid as long as the device is on an active service plan.[26]
- Operators typically transfer CDRs out of live stores (including switch databases, billing, mediation, and data warehouses) to tape/disk backup after three to six months. Once an object reaches tape backup, it is exceedingly difficult to obtain access as a third party, unless backed by a law enforcement need.
- Customer biographical information can be made available from the data warehouse via direct database transfer or CSV (comma separated value file) extract. The operator usually defines how long such information is retained.

It is relatively easier for third parties to gain access to operator datasets from downstream IT systems as opposed to upstream network, radio, and switching elements. Core operator systems are fairly standard across countries, but it is difficult to convince operators to grant third parties permission to access data from network and switching elements due to technical, commercial, and regulatory constraints. Data warehouse and billing/CRM systems are freer from such constraints and thus relatively easier to access.

The potential for successful collaboration is higher when operators are approached with requests for specific datasets, a defined use case, and an understanding of the commercial implications of the exercise.

3.6. Understanding and Interpreting Different Kinds of Mobile Data

The relevance of different network and back office systems to a research project depends on the nature of the information needed. Operators generate, process, and store data in vast quantities and across a diverse range of systems, subjects, and uses. Just as they rely on different mobile data sets to enable

[26] These are representative examples. IMSI is also available from multiple operator and Value Added Services (VAS) databases.

their various services and operational processes, so too can different mobile datasets help answer a variety of social development questions (Figure 18).

Figure 18 – Data Needs and Their Key Associated Mobile Systems

	Key Systems	Considerations
1. Location and Movement Data	**Core Network/ Switching System**	• Location area code/cell site ID is a passive update that is updated in real time, not stored as a standard data set, and is available only temporarily
	Handset	• Application and browser data are active updates that can help pinpoint location, but limited penetration in our target user groups hinders deeper analysis • Multiple approaches are available with smartphone-based location tracking
	Core Network/ Billing/Data Warehouse	• All three systems maintain CDRs that provide active updates on usage status • CDR contains location code, but data would only be available for users who generate a call/event on the network • CDRs are relatively easier to obtain, are a standard dataset, and provide historic data for analysis
2. Financial and Economic Data	**Prepaid System**	• Prepaid systems conduct continuous, real-time interrogation against available minutes and support real-time charging and checking of customer accounts • Pre-paid services are used by more than 99% of our target population • The voucher server contains records of each refill voucher, usage date and user, and value; historic analysis can show indication of users' income levels • The recharge server may store more granular recharge history per user
	Billing and CRM System/Data Warehouse	• The billing systems support aggregating service costs and tariff plans, producing a monthly bill and tracking payments from customers • The CRM keeps information related to products, services and customers and is responsible for tracking orders from creation till completion; the customer master and associated data stores provide information on users' credit limit, invoice/payment history, etc. • The data warehouse or business intelligence system combines data from across operator stores
	Core Network/ Switching System	• IMEI numbers from the CDR (voice, SMS, data) would help identify the model of the phone used, refresh rates, etc. and, therefore, would link to income level
3. Demographic Information	**Switching System**	• Operator system fragmentation may cause user profiles to be stored in multiple locations, such as the HLR in GSM/UMTS networks, in the HSS for IMS networks, and in SIP servers for the VoIP network
	CRM	• The customer master database stores the subscriber's name, age, gender, ID, etc., which are collected at point of sale • Different vendors provide customized solutions to operators, so database design, implementation, and possible data extraction would be on a case-by-case basis
4. Social/Browser Data	**Switching System/ Prepaid System/ Data Warehouse**	• CDRs enable analysis of social patterns emerging from call records and call metadata • Certain feature phone categories and smartphones provide potential IP addresses visited and social feeds based on CDR records, which can be mined to develop individual social graphs
5.	**Switching System/ Prepaid/Billing/**	• Usage events "light" up the network at different points, and ease of access vs. relative value trade-offs have to be made; individual patterns enable operators

Mobile Usage	**Data Warehouse**	to develop tailored products and services • The data warehouse provides the structures to improve visibility around churn, upsell, revenue leakage, customer satisfaction, etc. and is used as a primary source of management reporting
6. Social Sentiment and Trends	**Switching System**	• Anonymized analysis of CDRs, data mining techniques (e.g., sentiment analysis) applied on text messages etc. provide insights into the prevailing social sentiment among local populations
	CRM/VAS Stores	• User purchase of VAS services, and their periodicity, price, etc., helps understand trends such as culture and entertainment and also helps develop personal profiles
7. Diagnostic/ Ambient Conditions	**Radio Network/ Switching System**	• Analysis of signal propagation paths and interference leads to understanding ambient ecological conditions
	Handset (Smartphone)	• Device microphone, barometer, accelerometer, RF signal strength, etc. help to identify ambient conditions

As device functionality evolves and next-generation networks expand their reach, richer and more representative datasets will become available for analysis, enabling development professionals and researchers to explore a wider and more complex range of issues. The increased functionality of smartphones will result in richer mobile data generation that can be captured by multiple actors—not just operators—as other companies that provide browsers, smartphone operating systems, devices, content delivery services, advertising platforms, applications, and more also aggregate usage data.

Mobile data analysis allows researchers to derive mobility and location from cell tower connections, social networks from calling patterns, sentiments from text analysis, economic information from purchase behavior and usage of services and hardware, and even more when these data are combined. In addition, mobile data comes with a host of advantages because it is:

- Created for free as a byproduct of mobile usage
- Able to capture characteristics from a large part of the population, even extending to the rural poor who are currently “invisible” in most digital datasets
- Relatively high resolution in temporal and spatial dimensions (compared to other data tracking population movement and location)
- Created and accessible in real time (at least conceptually)
- A standardized dataset that allows for scaled analysis to scale to other contexts and countries

With billions of low-income people in developing countries adopting phones, combined with recent advancements in big data storage and analytic tools, major opportunities in both the commercial and development spheres are arising. The contours and scope of these opportunities are still coming into focus, but this paper begins to identify some of the special features of this data as well as some of the initial pilots and successes that help define it.

4. Present and Possible Applications

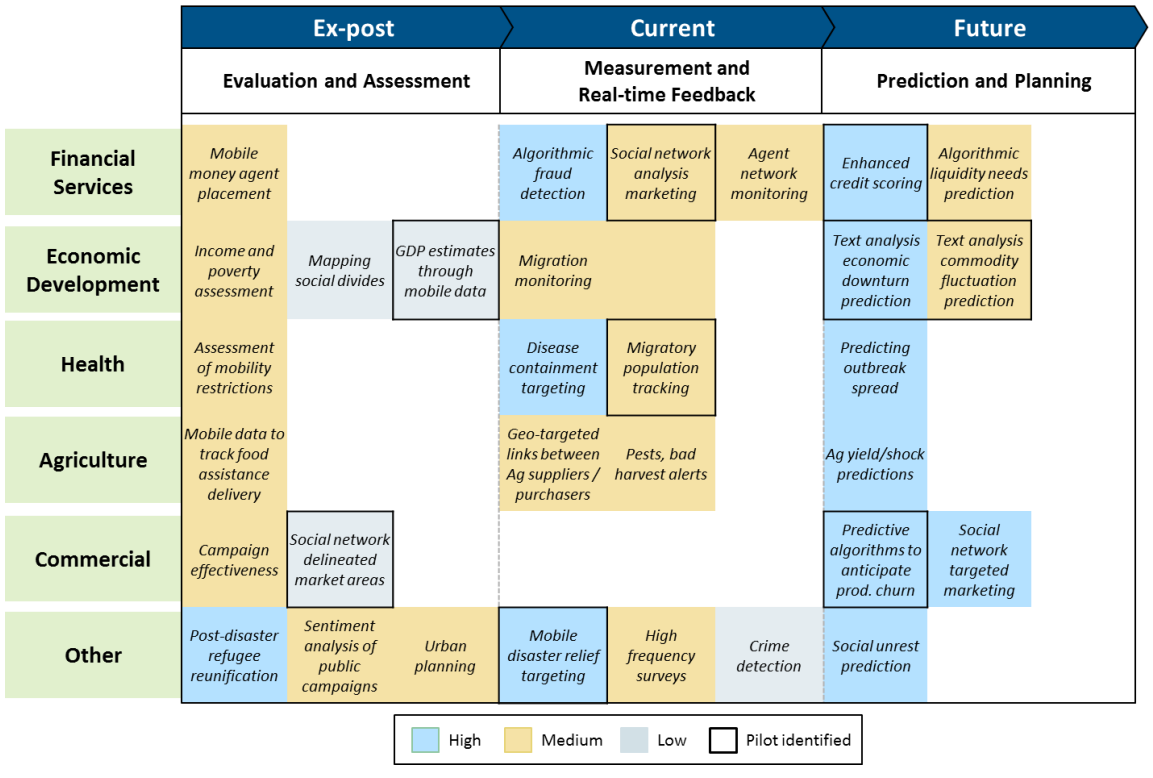
4.1. Potential Insights from Mobile Data

Mobile data offers a view of an individual's behaviors in a low-cost, high-resolution, real-time way. This provides tremendous opportunities for creative uses in development programs.

Figure 19 depicts the three primary types of analyses, and some of their applications: Ex-post – Evaluation and Assessment (e.g., estimating local wealth via past mobile phone activity); Current – Measurement and Real-Time Feedback (e.g., tracking population movements to understand where to deploy aid workers); and Future – Prediction and Planning (e.g., predicting where and when liquidity is needed along a mobile money agent's network) In general, the more predictive the analytics can be, the higher impact the analysis will have, but there are high impact applications for development and commercial practitioners across this spectrum.

To date, most mobile data for development efforts have focused on public health or emergency services initiatives. On the commercial side, the focus has been on customer segmentation for better-targeted marketing and churn reduction. One area where development and business interests meet is in the provision of financial services for the poor, since this is a social good that benefits significant populations in the developing world while also boosting revenues for banks, mobile operators, and other service providers.

Figure 19 – Areas of Highest Potential Impact across Sectors



Source: Cartesian, BCG

Note: "Other" category may include Emergency response, Mobility Planning, Law Enforcement, etc.

4.2. Mobile Operators Are Beginning to Exploit Mobile Data

Operators are increasingly exploring big data analytics to improve operations, develop new products and services, and generate more revenue. At the same time, greater numbers of software vendors and other big data analytics specialists are developing more effective, real-time, and sophisticated techniques and tools for capturing the full potential of operators' vast data stores.

Two overarching models are emerging, which are not mutually exclusive:

1. Driving internal capabilities: In this model, the operator uses big data analytics to drive operational improvements, develop better products and services inside its core businesses, and deliver a differentiated customer experience. Mobile network operators have primarily applied this model to develop better approaches to network optimization, customer retention, and churn reduction. For example, an operator in Rwanda and Ghana worked with third-party analytics consultants on anonymized subscriber data to conduct social network analysis to develop a predictive model that identified potential mobile money users based on their communities.[27]

2. Creating new products and services: Many different products or services could be developed relying on the platform of a mobile operator's dataset. For example, operators could sell insights to third parties (e.g., in the US, Verizon Precision Market Insights offers measurement solutions for a varied clientele, including media owners, advertisers, and venue owners.) Similarly, AirSage captures signaling data, CDRs, and other network traffic from operators, anonymizes and aggregates it, and provides insights to third parties.

Operators often engage in both models, where the typical evolutionary path is to begin by focusing on driving internal capabilities, and eventually expand efforts to include external opportunities as well.

Even in developed markets, most operators are only in the early stages of exploring the possibility of newly emerging analytical tools, with the majority of applications in the early phases of test and deployment. Facing the challenge of a maturing core business, many operators view monetization of their vast amounts of data as a key growth opportunity but are daunted by the task of managing and extracting value from the data.

[27] Primary research (vendor interviews).

The following case studies provide examples of an internal and an external data monetization:

Internal Data Monetization

Applying Big Data Insights to Enhance VAS Revenue

One major Indian carrier was trying to sell ringback tones (RBT) from its content catalog, but the average revenue per user (ARPU) had stagnated.

Subscribers could purchase an RBT by calling a special number to listen to five random songs, then pressing * to select one. If they didn't hear anything they liked, they had to wait and call again later to hear another five random songs.

The carrier hypothesized that the excessive amount of time it typically took subscribers to hear an RBT they liked was leading many to give up before purchasing.

To deal with the problem, the carrier deployed an analytics platform that used machine learning techniques and rules to analyze subscribers' demographics and individual consumption patterns to produce personalized RBT selections more likely to appeal to them.

This creative use of data analytics resulted in an increase in RBT revenue of 123 percent in just 45 days.

External Data Monetization

Using a Third Party to Derive Insights for Sale to Others

AirSage is a U.S.-based company that analyzes the wireless signaling data, CDRs, and other network traffic of partner operators, including two of the top three U.S. carriers, to generate location-based insights in order to better track people's movements for transportation planning purposes.

The firm's technology platform captures data directly from operator servers, anonymizes and aggregates it, and then sends it to a cloud-based platform for analysis.

This collection and analysis of real-time mobile signals produces more than 15 billion anonymous locations every day, which can be used by transportation professionals to create models for things such as toll road projects, reducing traffic congestion, and producing air quality studies.

This high-tech methodology takes the place of traditional methods of gathering data, such as household travel, vehicle intercept, and license plate surveys, which are expensive, laborious, and not always accurate.

Source: Company websites and interviews.

4.3. Current Uses of Mobile Data in Development Programs

By leveraging rich operator datasets and state-of-the-art analytic techniques, mobile data can help address a wide range of development needs across finance, agriculture, health, education and other spheres.

Much like the commercial sphere, the development world is only just beginning to understand the full potential of this type of data. Figure 20 below identifies some of the numerous examples of collaborations between operators and researchers who analyze CDR datasets to provide a window into the activities of a population. For example, the multinational operator Orange organized a D4D (Data for Development) Challenge that encouraged researchers to explore development applications using a modified (to protect privacy) set of Orange's CDR data on Ivory Coast subscribers. The result was a widely praised competition with over 80 research entries from leading academics and practitioners that showcased a wide variety of uses for this data.

While there have been a number of interesting research collaborations and some promising proof of concept studies, no significant program has yet been brought to repeatable scale leveraging mobile operator data for social good purposes.

Figure 20 – Pilot Program Uses of Mobile Data for Development

Program Area	Demonstrated Use Cases	Example
Financial Services	Mining social network information to determine optimal agent location	MTN
	Monitoring personal budget and expenses for developing credit profiles	inVenture
	Predicting purchasing intentions of individuals from real-time location tracking and regional trends	verizon
	Generating financial profiles of poor people with no access to traditional financial services to offer banking services	Safaricom
Economic Development	Monitoring and modeling population movement in the wake of natural or economic shocks	Digicel haiti
Health	Linking human mobility and connectivity patterns with spatial HIV distribution	
	Providing epidemic surveillance by mobile text messaging services	telenor group
	Launching targeted disease containment and information strategies in country-wide epidemics	
Agriculture	Developing financial identities with recharge history and subscriber data to offer microinsurance to protect farmers against drought	mobipay
Commercial	Achieving marketing efficiency by measuring campaign effectiveness	airtel
Urban Planning	Analyzing cell phone data to study intercity travels/traffic to better plan and reroute public transportation	orange
	Conduction origin-destination analysis for transportation planning such as new metro line and multimodal route coordination	IBM

Note: "Other" category may include Emergency response, Mobility Planning, Law Enforcement, etc.

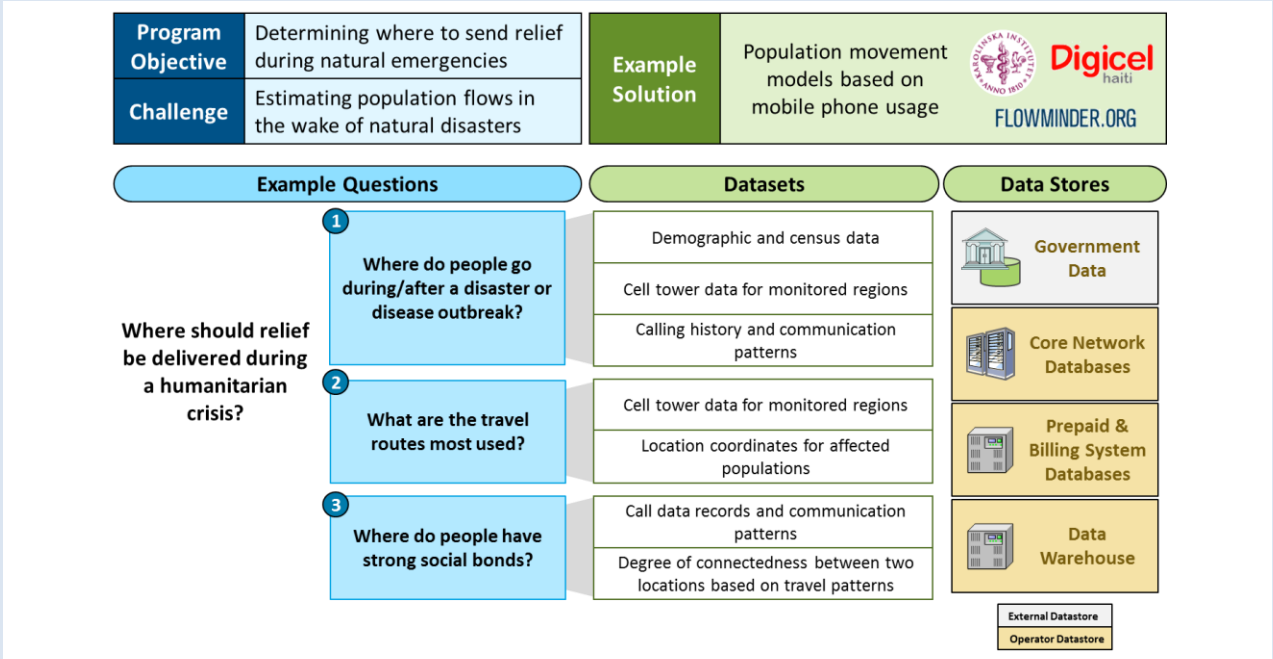
The use cases identified in Figure 20 demonstrate the scope of what is already possible with mobile data. While they have been proven academically or with a limited pilot, they have yet to be developed for systematic, large-scale use. Here we explore two use cases in greater depth.

Use Case 1: Disaster Relief

Using mobile data to estimate population flows in the wake of natural disasters and emergencies to determine where to send relief.

A natural disaster, conflict, famine, or major epidemic often results in en masse migration of populations from the affected areas. A challenge faced by relief organizations is how to effectively model population movements in such emergencies so that relief efforts can be organized and more effectively deployed. As long as the mobile infrastructure has not been completely wiped out, mobile data can provide the information needed to estimate population movements in near real time, which can help practitioners optimize the distribution of aid and relief services (Figure 21).

Figure 21 – Use Case 1: Disaster Relief



Flowminder is a nonprofit entity based in Stockholm that functions as a clearinghouse for aggregating, analyzing, and disseminating mobile phone location data to NGOs and relief agencies during disaster relief and reconstruction efforts. After the Haitian earthquake of 2010, a team from Flowminder and researchers from several U.S. and Western European universities analyzed cell tower data from 2 million SIM cards linked to Haitian operator Digicel to estimate population flows in the wake of the earthquake and a subsequent cholera epidemic.[28] They found that those who left Port-au-Prince after the

[28] Flowminder.org, Bengtsson et al, Lua, Bengtsson, & Petter Holme

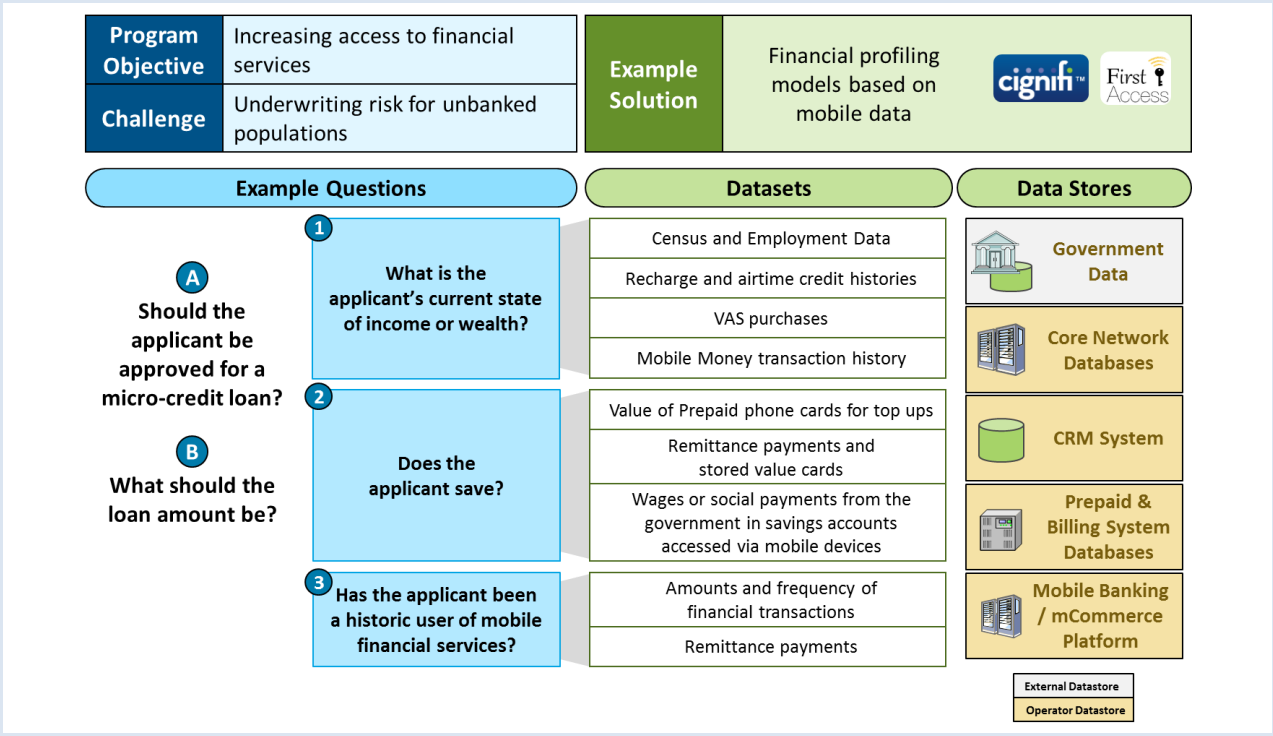
earthquake did not merely flee chaotically to the nearest "safe" zone, but instead had highly predictable travel patterns. Typically, survivors went to the location where they had spent the most recent Christmas and New Year's holidays, areas where they had strong social bonds. The cholera outbreak that began just months after the earthquake allowed researchers to validate their finding that people's travel patterns during more stable times predict their escape routes during crises. The Flowminder team's work provided strong evidence that estimating population movements during disasters and outbreaks using mobile data can be done rapidly and accurately.

Use Case 2: Financial Inclusion

Increasing access to financial services by using mobile data to generate financial profiles of unbanked persons.

Proponents of financial inclusion are beginning to see mobile data as an excellent way to build financial profiles of people who lack a conventionally documented financial history. Poorer, unbanked people have little to no record of past borrowing behavior and volatile income and expenditure patterns, making it difficult for banks and others to provide them with financial services such as savings products and access to credit. As a remedy, an individual's mobile usage data can provide proxy indicators, such as airtime usage, top-up history, mobile transaction data, and P2P transactions, to create an alternative financial profile[29] (Figure 22).

Figure 22 – Use Case 2: Financial Profiles for the Unbanked



[29] GSMA, Operator interviews, Cartesian.

The alternative profiling models already in use suggest the wealth of information embedded in mobile subscriber data.

A number of operators and banks have already begun to offer financial products that rely on mobile data indicators. In Kenya, for example, Safaricom and Airtel have formed partnerships with financial institutions to expand mobile money services to include short-term and longer-term microcredit provisioning. A number of aggregators, analytics services firms, and financial services specialists have developed mobile data-based automated profiling models and related services.

4.4. Future Opportunities to Leverage Mobile Data

The range of potential mobile data for development use cases goes well beyond the pilots that have been explored to date. Conversations with operators and researchers revealed a number of high potential applications which have yet to be tried. Figure 24 previews only a few of these specific possibilities.

Figure 24 – Illustrative Mobile Data-Driven Use Cases

Program Area	Potential Use Cases
 **Financial Services**	Using machine learning algorithms to predict liquidity needs across the agent network
	Leverage social network information to target marketing offers driving up-take
 **Economic Development**	Developing indicators of wealth, economic diversity and population segregation
	Developing ubiquitous sensing for mapping poverty in developing countries
	Understanding interactions between different ethnic, or socio-economic groups
 **Health**	Using location identifiers to send SMS or voice messages to residents in specific areas to warn them of epidemics or other health related risks
	Identifying immunization coverage rates (i.e., number vaccinated/ total population) using mobile phone data to estimate population in different regions
 **Other**	Detecting unexpected changes in weather quickly, by monitoring changes in cell patterns to develop an early warning system
	Monitoring rain patterns by analyzing background noise captured in voice calls
	Using smartphone battery temperature to determine changes in atmosphere/environment
	Combining call records and road data for strategic disaster response planning
	Developing traffic flow estimation models using cellular data instead of fixed sensor infrastructure

Note: "Other" category may include Emergency response, Mobility Planning, Law Enforcement, etc.

4.5. Should Mobile Data for Philanthropic Use Be Free?

One issue that often comes up when discussing the use of mobile data for development purposes is whether or not operators should be allowed to charge for the data they share. While many development use cases will no doubt justify some reasonable payment for access to data, there are also cases where the data should be shared for free.

The United Nations Global Pulse has put forward the idea of "data philanthropy," where operators would have a duty to share data for certain limited uses when the public good is urgent and clear. Global Pulse argues that these cases actually make business sense for a number of reasons:

- First, companies should want the best for their clients, if only because when their clients do better, they can afford more mobile services. In many cases, mobile data may hold clues to

upcoming problems, from disease outbreaks to agricultural crop failures. Global Pulse points to the following example:

"Imagine you are CEO of that company, and you have just completed construction of a number of costly new cell towers in a region that appears to be a promising market. Unbeknownst to policy makers, many in this community are being affected by an on-going, low-level food crisis. By the time this becomes public knowledge, your new customers are no longer able to afford your services."[30]

- Second, public backlash from refusing to share data could be significant, while, the goodwill generated among the non-profit and public sectors when sharing data could be of significant value. Recent blog discussions and press articles have picked up on the idea.[31]

Future debates will have to work out when data should be shared freely. Post-disaster scenarios seem to top the list, as do predictions of major economic shocks or disease outbreaks, but there are other uses where the mandate to share is less clear (e.g., ongoing monitoring of food prices).

[30] http://www.unglobalpulse.org/blog/data-philanthropy-public-private-sector-data-sharing-global-resilience (2011).
[31] http://www.unglobalpulse.org/data-philanthropy-where-are-we-now (2013).

5. Regulatory Landscape and Data Privacy Considerations

5.1. Regulatory Regimes Are Becoming Clearer and More Standardized

Though many of our target countries currently lack clear or comprehensive policies for mobile data protection, regulations are evolving rapidly as regulators attempt to keep laws relevant to advancing technologies, new national and personal security threats, and changing perceptions of privacy.[32] Due to wide variation in data protection laws, regulatory considerations have to be addressed on a per country basis. However, concerns about data privacy are becoming common across the globe, which is leading to laws that are more aligned across countries.

Regulatory controls on the potential use of mobile data can be divided into the laws and regulations that affect the *availability* of data and those that limit the *accessibility* of data. Two key regulatory structures that stipulate how data is collected and retained and thus affect the *availability* of mobile data are:

- *"Know Your Customer" (KYC) requirements:* These laws require operators to collect a particular set of identifying data (e.g., name, date of birth) about each of their subscribers before activating their mobile service accounts, so that every SIM card is associated with a uniquely identifiable person. Most target countries surveyed have similar standards for KYC. These regulations significantly enhance the quantity and validity of available mobile data.
- *Data retention requirements:* These laws require operators to store certain data on customers and their communications for a minimum (and sometimes maximum) period of time.[33] Countries around the world have increasingly adopted such mandates in order to support law enforcement efforts.

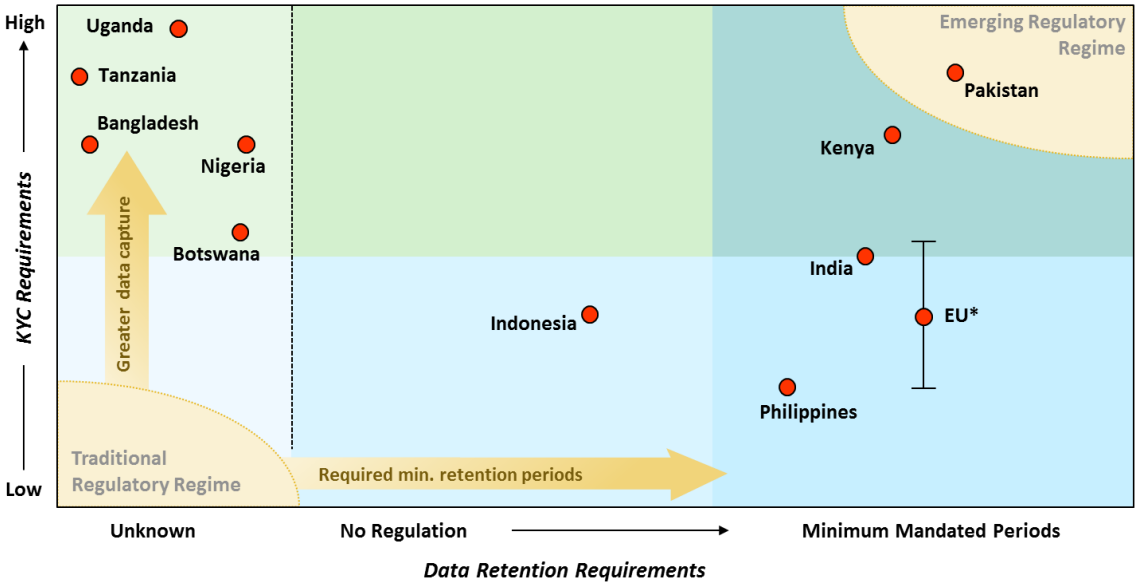
Conversely, data privacy and protection regulation limits the *accessibility* of data by establishing standards for the physical and technical protection of subscriber data and limiting when, how, and with whom data can be shared. Regulation relating to data privacy is also expanding in our target countries, and many operators have internal standards for data protection and privacy that match or exceed governmental requirements.

5.1.1. Data Capture and Retention Requirements

Recent trends in data capture and retention regulations have considerable implications for what data is available to operators. For example, regulators in many countries have begun introducing minimum standards for data capture (KYC) and retention among operators (Figure 25).

[32] Privacy International interviews.
[33] Telecoms.com.

Figure 25 – Landscape of Data Protection and Privacy Regulation



"Know Your Customer" Requirements Increase Data Availability

While KYC requirements have long been standard in banking services to limit the potential for criminal activity, equivalent rules for mobile services and mobile money services have been absent.[34] However, eight out of 10 of our target countries have recently mandated SIM card registration for prepaid customers, as KYC regulation for mobile service fast becomes the norm worldwide. SIM card registration has widened the scope of the user identification details that operators capture and store for each subscriber, which increases the overall availability of data.

The impact of SIM registration has already been significant—and sometimes controversial. In addition to requiring all new cards sold to be registered, most regulators have required operators to deactivate previously issued SIM cards still unregistered by a certain date. For example, after their deadline for registration passed, mobile network operators in Pakistan and Uganda had to block about 1.36 million and 1 million unregistered SIM cards respectively.[35]

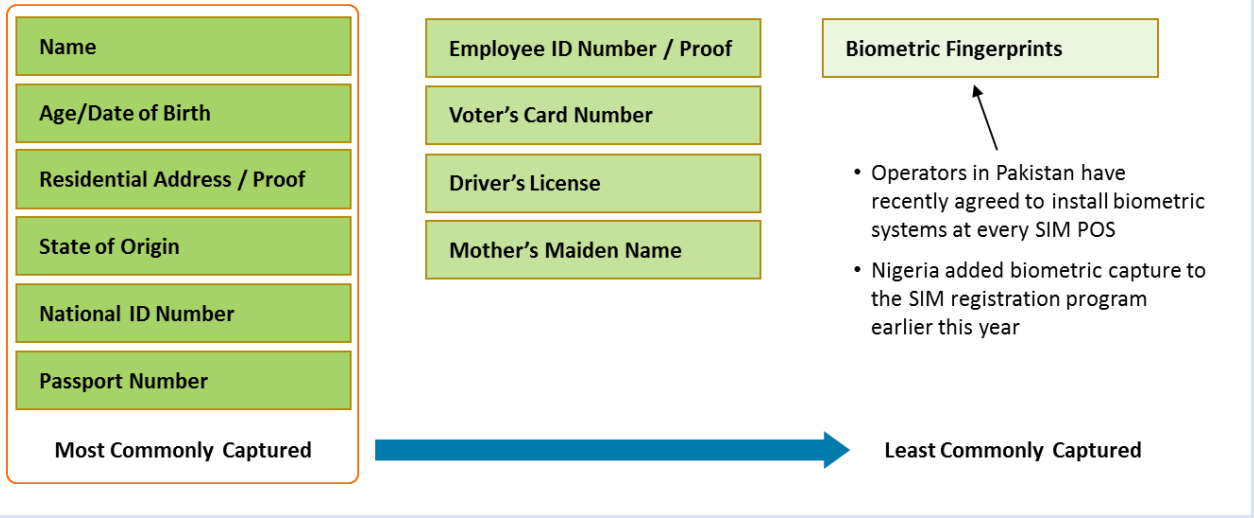
Despite these sweeping measures to block unregistered cards, KYC rules are still limited in their ability to create complete and accurate databases of subscribers' identities. Regulators have been unable to prevent individuals from buying and registering SIM cards in their own names, then giving or selling them to others. In an attempt to counteract this, Pakistan has mandated that each CNIC (Computerized National Identity Card, one per citizen) may be registered with no more than five SIM cards, and

[34] Ketkar, Shankar and Banwet, "Telecom KYC and Mobile Banking Regulation: An Exploratory Study," 2013.
[35] TeleGeography, Cellular News.

operators are required to check prospective subscribers' CNICs against a nationwide database to ensure they have not exceeded the quota before activating their new SIM cards.[36]

While the SIM card registration process is not yet standardized, many countries require operators to capture similar data due to KYC regulations (Figure 26).

Figure 26 – Know Your Customer (KYC) Data



Retention Laws Affect the Availability of Mobile Data

Data retention regimes vary based on how regulators balance the obligation to protect customers' privacy vs. the need to detect, prevent, and prosecute criminal activity. Increasingly, countries require service providers to retain data for some minimum length of time in order to ensure that it is available to the appropriate authorities should it be required in a criminal investigation. Other countries, including the United States, only mandate the "preservation of data," leaving it to each operator to decide what is a reasonable retention period.[37]

Operators define their data retention policies by balancing these regulatory needs with their business needs. As shown in Figure 27, different types of usage data are maintained for specified periods in operator systems. Core network and switching databases that capture "live data," such as location, do not normally serve as data stores. The VLR (visitor location register) for instance, typically overwrites mobile usage data as soon as a user moves into a different location area. CDRs such as those for voice calls and SMS text messages are also moved from the core switching or billing (rated) databases to the data warehouse, and subsequently to tape backup where they may be maintained for a number of years.

[36] Pakistan Telecom Authority.
[37] Center for Democracy and Technology, Washington, DC.

Figure 27 – Mandated Data Retention Periods

Country		Kenya	India (BSNL/MTNL)	Pakistan	Philippines	EU	US
Subscriber Data	Name						Unlimited
	Services Activated						Unlimited
Event Data	Call Detail Records		1 year	Min. 3 years	6 months	6 – 24 months	1 – 7 years
	Cell Towers Used		1 year		6 months	6 – 24 months	0 – 24 months
	SMS Detail Records		1 year		6 months	6 – 24 months	60 days – 7 years
	SMS Content				6 months from LEA request		0 – 90 days
	IP Session Info		3 years	90 days	6 months	6 – 24 months	0 – 1 year
	IP Destination Info		3 years	90 days	6 months	6 – 24 months	0 – 90 days
Financial Data	Bill Copies		3 years				0 – 7 years
	Payment History						Unlimited
	Recharge History		3 months				
	Mobile Money Transactions	7 years					

Retention Prohibited | Regulator Mandated Min. | Operator Defined Range | No / Unclear Policy

Source: PC Magazine, CDT, Country regulator websites, Gardner & Sonnenbergs/ Getting the Deal Through – Telecoms and Media 2012, CIS India

Some countries, such as Indonesia, lack any discernible rules for operators when it comes to data retention or at least have not publicized such requirements. Data retention regulation is currently being debated in some countries, with varying degrees of support and opposition. A survey of data retention laws in the countries highlighted here illustrates the paucity of specific regulation in the developing world:[38]

- **Bangladesh:** While the Telegraph Act 1885 allows law enforcement to intercept communications "in the interest of public safety," Bangladesh does not seem to have any regulation requiring data retention by service providers.
- **Indonesia:** Indonesia's Electronic Information and Transaction Act states that electronic system operators should retain data but does not specify any set retention period. Legal sources

[38] BTRC, BTRI, BTA, UCC.

suggest that though Indonesian operators and service providers have no specific obligation to retain customer data, they are required to share data for criminal investigations.[39]

- **Botswana:** The Communications Regulatory Authority Act (2012) does not obligate telecommunications licensees to retain subscriber data for minimum periods, nor does there seem to be any other regulation mandating such.
- **Tanzania:** Regulations under the Electronic and Postal Communications Act simply state that service provider licensees may collect and maintain information on individual consumers as reasonably required for their own purposes. No mandates are made on minimum retention periods.
- **Uganda:** The Posts and Telecommunications Act permits interception of telecommunications in public emergencies or for public safety, but does not indicate mandatory retention among service providers.

Apart from external regulatory requirements, the length of operators' data retention periods is also influenced by internal business considerations, including product and service development, better network traffic management, and fraud prevention.

5.1.2 Privacy Regulation and Data Accessibility

The expansion of regulation mandating various data privacy protections results from rising concerns about the risks to people's privacy and national security when data is exposed. Privacy protection frameworks are being developed to protect customers' civil rights and data privacy.[40]

The frameworks typically coalesce around several key principles captured in Figure 28.[41]

[39] Global Competition Review: Getting the Deal Through 2013- Indonesia.
[40] Privacy International.
[41] IAPP Privacy Academy Conference. PerspecSys, Perkins Coie.

Figure 28 – Key Principles for Data Accessibility

Data Ownership	Companies are only data custodians, while customers own and should have control over their data
Notice / Awareness	Customers should be made aware of what personal data is being collected and its use/purpose, and of any changes therein
Choice / Consent	Customers should be able to decide whether their data is collected, used, or shared to third-parties
Access / Participation	Companies should limit access to customer data in accordance with regulation, customer wishes, etc.
Integrity / Security	Companies should ensure that customer data is valid and that it is secure or safe from misuse
Enforcement / Redress	Regulators and other policy enforcers must ensure that customer data is protected

Note: These are generally accepted principles and guidelines that are used by data collectors in EU and North America

Source: IAPP Privacy Academy, 2013

http://www.perkinscoie.com/files/upload/lit_09_privacy_101_and_ethics_seminar_powerpoint.pdf

Personal data privacy protection has recently gained global attention and focus. Many countries explicitly address personal information privacy provisions in their constitutions, while most others mandate at least a basic right to privacy.[42] More than 80 countries have adopted data protection regulation beyond basic constitutional rights.[43] The EU's Data Protection Directive (1995) and subsequent revisions have shaped the laws of even countries outside Europe.[44] This has created a common legal framework that many countries are adopting in whole or in part, though there are still many local differences across markets.

Our target countries present a wide spectrum of maturity and completeness in data privacy and protection regulation (Figure 29). India and the Philippines provide the most complete legislative coverage of electronic data privacy and protection, though both still have limitations in some regards.[45] In Indonesia, Uganda, Nigeria, Pakistan, and Tanzania, personal data privacy and protection provisions exist but are typically dispersed through several separate laws or provide only incomplete protection of customers' right to data privacy.[46] One multinational operator in Pakistan described how it had to physically move the researchers it was partnering with in-country and co-locate servers within their local network in order to download, anonymize, and analyze data. Meanwhile, Bangladesh, Botswana, and Kenya solely address personal data privacy and protection through an overarching constitutional right to privacy with few to no specific statutory provisions. As of fall 2013, Kenya and Botswana are actively debating laws on data privacy and protection and freedom of data access.[47]

[42] Privacy International, Graham Greenleaf- Global Tables of Data Privacy Laws and Bills (2013).
[43] Electronic Privacy Information Center.
[44] McAfee- International Privacy and Data Protection Laws, European Digital Rights.
[45] BakerHostetler, TRAI, NTC.
[46] World Legal Information Institute, EPIC, Norton Rose Fulbright, DLA Piper.
[47] AptanTech, ITU.

Figure 29 – Privacy Regulations: Country-Specific Highlights

Country	Relevant Regulation	Regulatory Coverage	Comments
Bangladesh	• Bangladesh Telecommunications Act (2001) • Information and Communication Technology Act (2006)	☐ Security Practices ☐ Consent ☐ Informed Collection ☐ Limited use ☐ Access by data subject ☐ Limited data retention ☐ Mandatory privacy policy ☐ Limit on external transfer ☑ Court-ordered access by law agencies	• Right to privacy still nascent • Covered mainly under constitutional protection
India	• Information Technology Act 2000 (Amended 2008) • Information Technology Rules (2011)	☑ Security Practices ☑ Consent ☑ Informed Collection ☑ Limited use ☑ Access by data subject ☑ Limited data retention ☑ Mandatory privacy policy ☑ Limit on external transfer ☑ Court-ordered access by law agencies	• India's constitution does not specifically address a citizen's right to privacy, nor is there a statute protecting data in all possible settings
Indonesia	• Information and Electronic Transaction Act (IETA)	☑ Security Practices ☑ Consent ☐ Informed Collection ☐ Limited use ☐ Access by data subject ☑ Limited data retention ☐ Mandatory privacy policy ☑ Limit on external transfer ☑ Court-ordered access by law agencies	• Although the main driver for the IETA was to prohibit the viewing and sharing of pornography, it includes a high level protection against the use of electronic information without consent
Tanzania	• Electronic and Postal Comm. Act (2010) • Electronic and Postal Communications (Consumer Protection) Regulations (2011)	☑ Security Practices ☐ Consent ☐ Informed Collection ☐ Limited use ☑ Access by data subject ☐ Limited data retention ☐ Mandatory privacy policy ☑ Limit on external transfer ☑ Court-ordered access by law agencies	• In addition to these, Tanzania's constitution explicitly expresses a broad right to privacy, as well as to the respect and protection of private communications
Nigeria	• Nigeria Evidence Act • [Draft] Data Protection Act	☐ Security Practices ☑ Consent ☐ Informed Collection ☐ Limited use ☐ Access by data subject ☑ Limited data retention ☐ Mandatory privacy policy ☑ Limit on external transfer ☑ Court-ordered access by law agencies	• The Government has drafted a Data Protection Act that has been described as inadequate to protect citizens' privacy • E.g., data subjects are not given express right to obtain their own personal data
Pakistan	• Pakistan Penal Code • Pakistan Telecommunication (Re-organization) Act (1996) • Electronic Transaction Ordinance (2002)	☐ Security Practices ☑ Consent ☐ Informed Collection ☐ Limited use ☐ Access by data subject ☑ Limited data retention ☐ Mandatory privacy policy ☑ Limit on external transfer ☑ Court-ordered access by law agencies	• No current law fully protects the privacy of personal data • Regulators are currently considering a *Prevention of Electronic Crimes Bill* and an *Electronic Data Protection Act* (based on a subset of EU standards)
Philippines	• Data Privacy Act (2012) • Cybercrime Prevention Act of (2012)	☑ Security Practices ☑ Consent ☑ Informed Collection ☑ Limited use ☑ Access by data subject ☑ Limited data retention ☐ Mandatory privacy policy ☑ Limit on external transfer ☑ Court-ordered access by law agencies	• Recently passed comprehensive laws addressing emerging concerns on cybercrime and data privacy
Uganda	• Constitution • Access to Information Act (ATIA) • Communications Act	☑ Security Practices ☐ Consent ☐ Informed Collection ☐ Limited use ☑ Access by data subject ☐ Limited data retention ☐ Mandatory privacy policy ☑ Limit on external transfer ☑ Court-ordered access by law agencies	• The ATIA asserts citizens' right to access data held by state agencies, (except when affecting another's privacy or state sovereignty) but does not cover access to data held by private parties

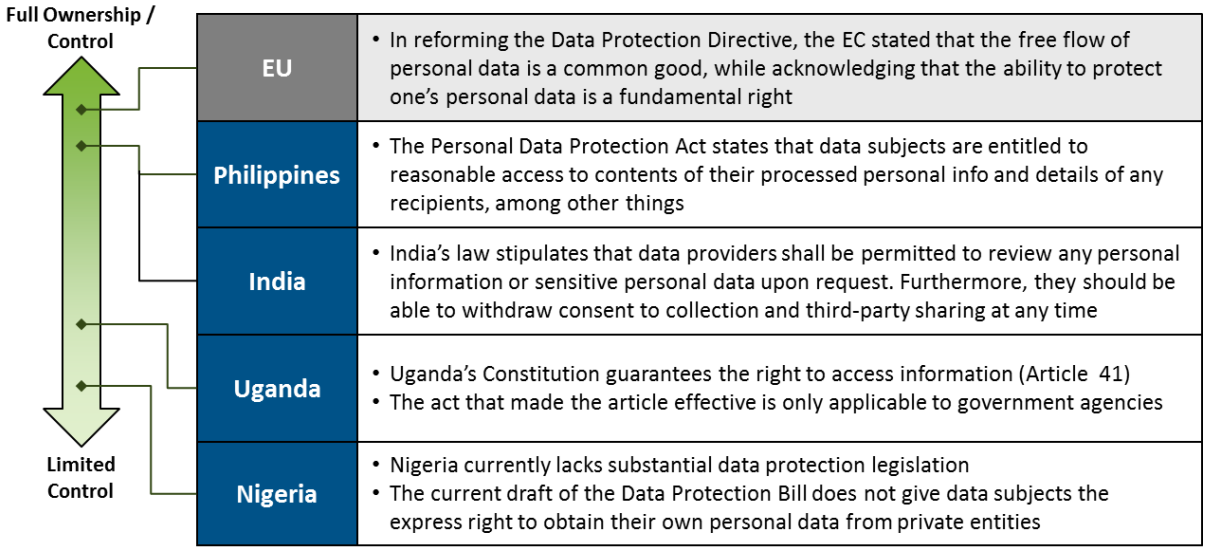
Note: Insights are based on primary interviews and secondary research and are not to be construed as a professional legal opinion

Source: Cartesian Primary Interviews, Privacy International, BakerHostetler, Article 19, World Legal Information Institute, EPIC, Norton Rose Fulbright, DLA Piper

Subscriber Data Ownership

Many privacy frameworks start from the principle that subscribers own their own data. Subscriber data ownership can apply to personally identifiable data such as address and name, and in some cases even to anonymized or aggregated data. In the case of identifiable data, subscribers' access to and control over their personal data varies from country to country. Some countries, such as Colombia, define subscribers as the owners of their data, while operators are defined as the custodians. Typically, laws in our selected countries do not explicitly state that subscribers are the owners and controllers of their personal data (Figure 30). Even when not required by law, most operators we spoke to indicated that they viewed themselves as the custodian of customer data not the owner and primarily accessed this mobile data for internal improvements and law enforcement compliance.

Figure 30 – Subscriber-Level Personal Data Ownership and Control



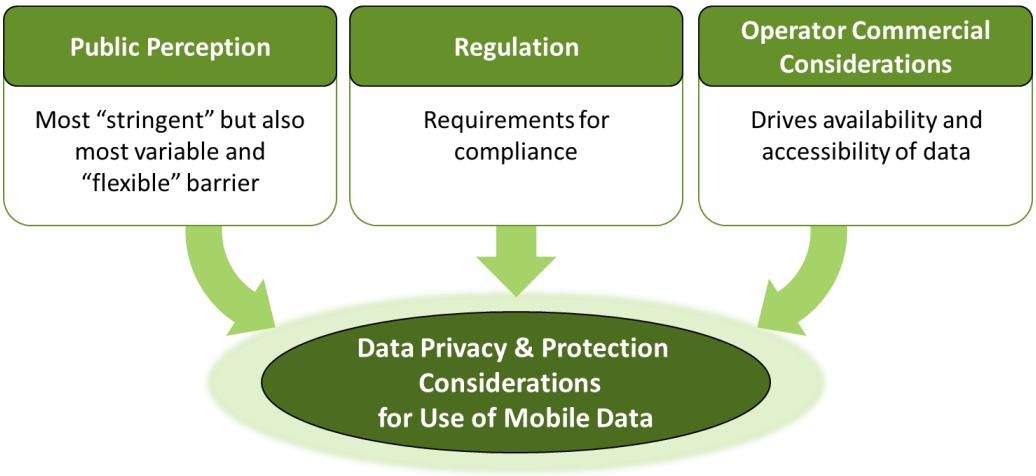
Source: Country Regulations

6. Considerations for Data Sharing

6.1. Sensitivities around Mobile Data Access

While mobile data can be valuable for a variety of development projects, those seeking access will need to consider the sensitive nature of this data. As Section 5 above makes clear, they will have to contend with the country's regulatory environment, since most governments set requirements on how to obtain, process, store, transfer, and safeguard data to protect individual privacy. Next, they will have to deal with the operator's own data protection policies, as well as the company's commercial considerations, which include its perception of the costs and benefits of collaborating, the reputational risks involved, and competitive market forces. Finally, with public perceptions of privacy evolving every day, those holding data must consider how the public will react to their use of this personal information (Figure 31).

Figure 31 – Data Privacy Context



Data anonymization (removing personally identifying information) goes some way toward mitigating potential risks and alleviating privacy concerns but can't be relied on to protect data fully because data can often be de-anonymized by triangulating against other data sources as described below. In cases where data that is not anonymized or aggregated must be used, and even in cases where it is, an opt-in/opt-out model can be implemented in order to obtain customer consent. These topics are covered in more detail below.

6.2. Commercial Sensitivities around Data Access

Operators' actions and attitudes to collaboration are determined, in the most part, by their commercial interests. As mentioned above, operators are increasingly interested in exploring different ways of monetizing mobile data via new business models. However, they must weigh each of these new monetization opportunities against the costs associated with the extraction and retention of additional data, the risk of jeopardizing customer relationships, and the risk of creating negative public perception. Additionally, when the commercial potential is unclear, operators can be hesitant to partner to explore new applications of the data for fear of losing commercial advantage.

Regulatory grey areas create uncertainty and may deter innovation

Public officials in some of our target countries have not yet been able to develop effective regulatory regimes, so operators in these markets have had to devise measures to overcome the lack of privacy standards and to harmonize data management practices within their global operations. Specifically, they have relied on three key risk mitigation strategies to fill the gap left by state regulators.[48] First, they sometimes self-impose industry standards, such as the ones issued in 2011 by the GSMA, a trade association of about 800 operators in more than 200 countries. GSMA's Mobile Privacy Principles serve as a foundation for developing guidelines or codes of conduct for all members. Second, multinational operators might voluntarily apply regulations propagated in their home countries in countries that lack such regimes. For example, Orange has implemented EU-type policies in Uganda. Finally, operators are careful to maintain close working relationships with regulators so they can stay ahead of changing regulations and clarify ambiguous expectations, among other considerations.

6.3. Public Opinion Plays a Role

In addition to regulatory and commercial concerns, researchers need to ensure they are sensitive to public perception. For example, when researchers at the Harvard School of Public Health and other institutions obtained Kenyan mobile data records to track the spread of malaria in 2012, they provoked controversy among Kenyans who had unknowingly contributed data to the study.[49] The researchers obtained anonymized CDRs for every call made and text message sent by 15 million Kenyan mobile phone subscribers between June 2008 and June 2009 and used this data to identify possible regions of infection origination for targeted health interventions. Despite the fact that standard precautions were taken to anonymize the data and assure its use was only for legitimate purposes, a number of local press articles, written in inflammatory language, caused significant public discussion regarding the researchers' methods and intents. This incident demonstrates that even with the best of intentions and adherence to rules, researchers need to proactively manage and anticipate public perceptions.

6.4. Approaches to Protecting Data Privacy

While regulatory issues, commercial concerns, and public perception challenges are real, researchers can develop a set of tools and guiding principles to mitigate their effect, including clearly articulated privacy policies. The United Nations Global Pulse's Privacy and Data Protection Principles can serve as a helpful example.[50] The issues that must be considered by researchers and development practitioners include:

- National and personal security
- Privacy and consent
- Fraud prevention
- Integrity of data and retention
- Informed consent
- Technical and administrative safeguards
- Notice/awareness of data use

[48] GSMA, Operator primary interviews.

[49] HSPH, TMC News, Kenya Daily Nation, Boston Globe, Technology Review.

[50] See General Assembly resolution 45/95 and http://www.unglobalpulse.org/privacy-and-data-protection

- Due diligence when selecting data or service provider partners

Data anonymization, controls on data access, clear communication, and thoughtful use of opt-in/opt-out models can help mitigate potential risks and alleviate customer concerns about privacy. In the long run, development practitioners can also play a role in getting ahead of future and present concerns by advocating for clear, consistent, and thorough regulation in the countries where they operate.

Key approaches to working with personally identifiable data

At the heart of these considerations is often the need to respect individual rights to privacy regarding personally identifiable information (PII). The EU Data Privacy Directive (DPD) is one of the more influential pieces of regulation relating to data privacy globally as it defines standards in Europe and has been copied in many other countries. The DPD defines PII as:

> *"'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity;"* – Article 2; Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995

Ultimately, there are four primary ways to adhere to data protection obligations when handling data (though the sufficiency of each of these needs to be determined in each jurisdiction and sometimes on a case by case basis):

1) *Waiver of rights:* Users may opt in to the service, allowing his or her data to be used for these applications. Not opting out is also permitted in some circumstances, though not generally considered to meet the standard of informed consent since many people don't realize they have the option to opt out.
2) *Statutory rights of access:* Rights of access to data for law enforcement and other activities are sometimes instantiated in law and can override the general right to privacy. Similarly, the DPD and other frameworks allow some room for uses of the data in the public interest or for scientific purposes.
3) *Anonymization:* Data is sufficiently anonymized so as to obscure any "personally identifiable information," which regulatory bodies and other standard setters agree is sufficient. In this case, usually the higher degree of anonymization, the greater the degradation of the fidelity and information content of the data.
4) *Access control:* Access to the data should be limited to only those people who are bound to use the data in appropriate ways. Researchers can get access to data when conducting research in the public interest with contractual obligations for non-disclosure and agreement to not de-anonymize data and/or use the data for purposes other than those stated at the outset of a project, which regulatory bodies and other standard setters agree is sufficient.

While 1) and 2) amount to cases where rights to privacy have been waived or don't apply, 3) and 4) are ways to protect the right of privacy while still working with data for the public interest or for scientific ends.

Anonymization and access control are used in conjunction to achieve the appropriate level of protection and data fidelity. For example, use of data only by select individuals who are contractually bound to use the data appropriately and have a high incentive to do so (e.g. scientists) might require more limited

anonymization applied to the data. Conversely, data that is shared with a wider audience, for example in a public data competition might need to be more coarsely anonymized and aggregated.[51]

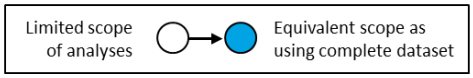
Interesting platforms, including AirSage's analytical platform, Palatnir's data management platform, or CryptDB developed from MIT all make attempts to create a hosting site for personally identifiable information, and address privacy concerns through appropriate access protocols. Going forward, new access technologies like these which can facilitate appropriate access to results of data analyses while still maintaining the anonymity of individual PII may open a wider range of use cases for data.

Anonymization Protects Identities

Anonymization techniques are used to prevent the identification of individuals through their personal data.[52] The operator will almost always complete some form of anonymization prior to sharing data with any third party. In deciding what anonymization techniques are appropriate, operators and researchers must find a balance between privacy needs and the usability of the data after key fields are removed or blurred. Figure 32 explains the trade-offs different approaches present. It should be noted that these approaches are not mutually exclusive and can be complementary.

Figure 32 – Anonymization Approaches

Approach	Suppression		Generalization		Perturbation		Permutation	
	Removing sensitive personal information or fields from data (e.g., name)		***Coarsening information into sets or clusters (e.g., location, time period)***		***Injecting noise into the data (e.g., salaries)***		***Switching sensitive associations between entities (e.g., gender)***	
Range of Possible Analyses		Broad range possible if removed fields were not necessary for analyses		Less granular data constrains the granularity of analyses		Broad range of analyses as the database should mirror original		Broad range possible if permutated fields are not necessary for analyses
Privacy Level *(Robustness)*		Individuals or groups can be identified when individual patterns are paired with other datasets		Depends on level of coarseness or clustering		Blurs whether specific fields are included in the database		Depends on the diversity of permutated field



Limits to Anonymization

While data anonymization can help address regulatory and privacy concerns, it is not foolproof. For example, even if CDRs are stripped of personal identifiers such as names and numbers, the identity of the person can often be revealed through other means, such as by combining with external datasets. Several such incidents have been reported. In some cases, individuals were linked to their health records

[51] In 2012 Orange sponsored a data competition (Orange D4D) where aggregated and anonymized data sets were shared with an over 80 research teams who submitted applications and signed agreements to treat the data appropriately.

[52] AT&T Labs, Sprint, Simon Fraser University, Princeton University.

because demographic information, such as age, gender, and postcode, is contained in both these records and in other data sets such as voter lists. In another example, researchers re-identified individuals in the Chicago homicide database[53] by linking it with the social security death index. And in another, a researcher cross-referenced voter lists with public record information from the Group Insurance Commission to identify the medical records of former Massachusetts Governor William Weld.[54]

As a result, data users have to understand that data is not either identifiable or not. Instead, there is a continuum of identifiability, as described by El Emam: on one end is data that is relatively easy to re-identify, and on the other is data that requires "considerable time, effort, cost, and skill to re-identify."[55] El Emam suggests coming up with a threshold on the continuum, wherein data that is above the threshold would be considered personal and identifiable, while data below is not. The goal would then be to define where that threshold sits. Sweeney puts forward a model to protect data from being re-identified, noting ways that the model can be attacked and ways those attacks can be thwarted.[56]

Opt-In and Opt-Out Consent Models Provide a Range of Solutions

Anonymization techniques allow for sharing of data by removing sensitive information, but suffer the drawbacks cited above. In some cases, there may be a desire by a mobile subscriber to share personally identifiable information and other sensitive elements of the data. Typically this can only be done with subscriber permission.

Under an opt-in model, operators actively seek customer consent prior to sharing data. Under an opt-out model, such consent is implied because customers are given the opportunity to remove themselves from data sharing and collection programs with an opt-out option.[57]

The likelihood of consent under an opt-in model, which requires active agreement from the user, varies by business model and the benefit users perceive in participating. Therefore, this model may yield less data, as some subscribers may think it is too costly to share personal data. Opt-in models are increasingly being required for third-party data sharing. For example, starting in 2014 Indonesia will switch to a regulatory regime that requires opt-in for all data sharing with third parties.[58]

An opt-out model, with its requirement of only passive consent, offers a high likelihood of consent because it is the default setting. However, by using this model, operators run a greater risk in angering subscribers who were not aware that their data was being captured and shared.

Operators often decide between these two options based on the nature of the data that is being collected as well as how it is used and shared. For example, some providers in the Philippines assume subscriber consent to sharing aggregated data with third parties and affiliates while providing an avenue for opting out. In contrast, Verizon Selects seeks opt-in consent prior to sharing data with third parties. This marketing analytics service from Verizon Wireless in the US segments customers based on usage, location, web browsing, demographic, and interest data; develops insights about certain segments; and

[53] Ochoa S, Rasmussen J, Robson C and Salib M. "Reidentification of Individuals in Chicago's Homicide Database: A Technical and Legal Study," 2001.
[54] Sweeney L. "K-Anonymity: A Model for Protecting Privacy," International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5): 557-570, 2002.
[55] El Emam K. "Risk-Based De-Identification of Health Data, Security & Privacy," IEEE 8(3) 64, 67, 2010.
[56] Sweeney, 2002.
[57] Bouckaert and Degryse. "Opt In Versus Opt Out: A Free-Entry Analysis of Privacy Policies," Operator privacy policies, 2006.
[58] Cartesian primary interviews.

allows third parties to individually target subscribers in desired segments with promotional messages. In exchange for opting in, subscribers are awarded a coupon or a similar reward from popular retailers.

7. CONCLUSION

Mobile data has enormous potential to support development efforts and through this to improve the lives of poor people around the world. Researchers have demonstrated the power of these data sets to improve financial services, support disaster relief, and to track and prevent the spread of infectious disease, to name just a few of the possible opportunities. Mobile data offers larger and more representative samples, in near real-time, and at far lower costs than alternative means of data gathering. Indeed we believe the opportunities to leverage these data sets for development goals are only starting to be explored.

While the examples are encouraging, too often these projects are trials or proofs of concept and have not been repeated or broadened to the fullest extent possible. There is an opportunity to learn from the best examples that have been demonstrated to develop a foundation for broader use of mobile data through a range of potential activities:

- Establishing mechanisms to coordinate and collaborate between researchers and mobile network operators to identify and execute on development opportunities.
- Building a structured and managed framework for submitting and approving research requests in ways that protect mobile users' personal information and the commercial objectives of service providers.
- Allowing for data to be combined with third party and government data sets to enrich and accelerate research efforts.
- Developing clear standards and nuanced best practices for handling privacy issues.
- Creating a technical architecture to automate data aggregation and re-distribution processes, possibly through a trusted clearinghouse or set of approved and controlled data repositories.
- Providing forums to replicate and scale successful development examples beyond the trial or proof of concept phase.

In order to leverage the power of mobile data sets, trusted relationships between operators and researchers must be formed. Data must be made available and accessible, which will require both regulatory solutions on the part of governments and technological solutions on the part of operators. Complex issues of privacy must be confronted, including the very volatile relationship the public has with the concept of having their personal information shared.

In the end, the success of using mobile data to enhance development initiatives will depend on the cooperation of all parties involved: governments, mobile operators, researchers, and mobile phone users. Once all of these entities fully understand the power of this data and recognize its ability to solve difficult problems and improve people's lives, the opportunities for development will only be bound by our collective imaginations.

Glossary

2G	A set of standards for “second generation” wireless, first to offer an inexpensive implementation of SMS (texting)
3G	A set of standards for “third generation” wireless, expected to deliver quality multimedia to mobile devices by way of faster and easier wireless communications as well as “anytime, anywhere” services
4G	A set of standards for “fourth generation” wireless, expected to securely provide mobile service users with bandwidth higher than 100 Mbps, enough to support high-quality streaming multimedia content; includes LTE and WiMAX
A Number	The calling number in a voice call
API	Application programming interface, a set of functions or routines that specifies how different software components should interact with each other
APN	Name of a gateway between a GPRS (or 3G, etc.) mobile network and another network (Internet, private customer network, etc.)
B Number	The called number in a voice call
Backhaul	Network links that support the process of getting data to the core (backbone)
BoP	“Bottom of the pyramid,” the largest but poorest socio-economic group; the 4 billion people who live on less than $2.50 USD per day
CC	Country code, a component of MSISDN
CDR	Call detail record, a computer record created by a telephone exchange covering the details of a phone call established through the telephone exchange, including an automated record of the length of each telephone call
Cell Site	Physical location in a cellular network where antennas and communications equipment are placed
Clickstream	A record of what computer users are clicking while browsing the Web
CRM	Customer relationship management, a strategy widely used by companies and organizations to record and manage their overall data and interactions with current, past, and potential customers
CSV	Comma separated values, a CSV file is a set of database rows and columns stored in a text file; primarily used to transport data between two databases of different formats through a computer program
DWH	Data warehouse, a unified database that holds all the business information of an organization and makes it accessible all across the company
EDR	Event data record, records created to record a transaction and affixed with a timestamp, e.g., a CDR
FAC	Final assembly code, a component of IMEI; a manufacturer-specific code indicating the location of a device's construction
GPRS	General packet radio service, a mobile data service on the 2G and 3G cellular communication system's global system for mobile communications; typically charged based on volume of data transferred
GSM	Global system for mobile communication, a second generation (2G) standard for mobile networks
HLR	Home location register, a database containing pertinent data regarding subscribers authorized to use a GSM network; the HLR serves as the main source of recent location information because it is updated each time the SIM transfers into another location area
HSS	Home Subscriber Server
IMEI	International mobile station equipment identity, a unique 15 to 17 digit number identifying a GSM device
IMSI	International mobile subscriber identity, a unique number, usually 15 digits, identifying a GSM or UMTS subscriber
ITU	International Telecommunication Union, the U.N. agency responsible for issues concerning information and communication technologies; assists in the development and coordination of worldwide technical standards, among other responsibilities
KYC	Know your customer, due diligence activities that financial institutions and others must perform to ascertain relevant information from their clients for the purpose of doing business with them; intended to

	prevent identity theft and other crimes
MCC	Mobile country code, component of IMSI
MMS	Multimedia message service, a mobile content exchange mechanism that allows users to transmit and receive videos, images, ringtones, and text files
MNC	Mobile network code, component of IMSI
MSC	Mobile switching center (MSC), the centerpiece of a network switching subsystem (NSS); mostly associated with communications switching functions such as call set-up, release, and routing
MSIN	Mobile subscriber identification number, the 10-digit unique number that a wireless carrier uses to identify a mobile phone; the last part of the international mobile subscriber identity (IMSI)
MSISDN	Mobile station international subscriber directory number, the telephone number of the subscriber identity module (SIM) card displayed on mobile or cellular phones; uniquely classifies a subscription in the GSM or UMTS networks
MSRN	Mobile station roaming number, a telephone number used to route telephone calls in a mobile network from a GMSC (gateway mobile switching center) to the target MSC
NDC	National destination code, a component of MSISDN
SDP	Service delivery platform, a platform that provides a structure for service delivery, including controls for service sessions and protocols for service use
SIP	Session initiation protocol, a text-based signaling protocol that establishes Internet protocol (IP) network sessions at the application layer
SMS	Short message service, the most basic communications technology for mobile data transfer; characterized by the exchange of short alphanumeric text messages between digital line and mobile device
SMSC	Short message service center; the part of a wireless network that manages SMS operations, which includes storing, routing, and forwarding inbound short messages to their desired endpoints
SN	Subscriber number, a component of MSISDN and MSRN
SNR	Serial number, a component of IMEI
Switching	A networking communications method by which data is transmitted by being passed along the network from node to node
TAC	Type approval code, a component of IMEI
Trunk Group	Connect switching centers; can be used to learn data and voice traffic patterns
UMTS	Universal mobile telecommunications system, an umbrella term that encompasses the third generation (3G) radio technologies developed for a 3G mobile system based on evolved GSM core networks as well as the radio access technologies associated with them
VAS	Value-added service, any service beyond core services (standard voice calls and fax transmissions) available at little or no cost; for mobile phones, technologies like SMS, MMS and data access were historically usually considered value-added services, but in recent years have become core services
VCC	Visitor country code, a component of MSRN
VLR	Visitor location register, a database that contains information about the subscribers roaming within a mobile switching center's (MSC) location area
VNDC	Visitor national destination code, a component of MSRN
VoIP	Voice over Internet protocol, a technology used for delivering telephone calls from a source to a destination using IP (Internet protocol)
XML	Extensible markup language, a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable

Appendix: Additional Mobile Data-Driven Use Cases

Financial Services	CGAP study on social network analysis (2013) demonstrates higher marketing return on investment expected from those customers who are technology leaders and who also have a high number of connections with active mobile money users.
	Analysis of consumer mobile data to create individual financial profiles that allow poor people in emerging markets without conventional financial documentation to access banking and lending services. Examples include M-Shwari service from Safaricom in Kenya, and solutions from First Access, Wonga, etc.
	Providing underwriting services using machine learning and large-scale big data analysis (e.g., ZestFinance)
Economic Development	UN Global Pulse and PriceStats conducted real-time tracking and analysis of commodity prices. In the 2013 study, the relationship between web-extracted prices and official statistics on food prices (i.e. bread) proved to be closely correlated, allowing for price forecasting and additional real-time indicators of inflation activity.
	UN Global Pulse and SAS conducted a study on using social media as an early indicator of an unemployment hike, by performing sentiment analysis to categorize the mood of online conversations.
	Researchers understanding socio-economic indicators in UK using CDRs (Eagle, et al., 2010) found that more diverse social ties correlate with better access to social and economic opportunities.
Health	Migratory population tracking, e.g., quantifying the impact of human mobility on malaria. In this study, Buckee, Nathan, et al. (2012) used spatially explicit mobile phone data and malaria prevalence information from Kenya to identify the dynamics of human carriers that drive parasite importation between regions. The analysis helped identify the sources and sinks of imported infections due to human travel and locate high-risk sites of parasite importation.
Commercial	A study conducted at MIT's Senseable City Lab, "Delineating geographical regions with networks of human interactions in an extensive set of countries," used telecommunications data to demonstrate social, civic, and commercial interactions (or lack thereof) between different regions in selected countries
	Use of predictive analytics by mobile operators to anticipate customer and product churn, so as to be able to better target at-risk customers with retention programs.
Other	Mobile disaster relief targeting: Researchers have demonstrated improved response mechanisms to disasters and outbreaks by tracking population movements with mobile phone network data (e.g., Bengtsson, et al. "A Post-Earthquake Geospatial Study in Haiti")

Two Financial Center
60 South Street, Suite 820
Boston, MA 02111
+1 617 999.1000

www.cartesian.com

PO Box 23350
Seattle, WA 98102
+1 206 709.3100

www.gatesfoundation.org