

Human Mobility Data from Anonymized Mobile Phone Call Data Records:

Use cases and Applications for Social Good

Data Scientists Network https://www.datasciencenigeria.org/ November 28, 2022

Contents

2	Roadmap	6
2.1	Stakeholders	7
3	Scoping the Data Landscape	9
3.1	Privacy	14
3.2	Leveraging AI and Machine Learning for CDR Data	15
4	Use Cases: Organizational/Infrastructural Optimization using CDR	18
4.1	Optimizing Public Transportation System Using Subscribers Density and Point of Interest	19
4.2	Precise Waste Management Using Call Data Records and Satellite Imagery	20
4.3	Digital Inclusion and Mobile Penetration in Nigeria Using CDR data	21
4.4	Police and Security Alertness using Aggregated Call Data Records	22
4.5	Optimizing Electricity Allocation By the Dynamics of People's Movement	23
4.6	Traffic Analysis and Prediction using CDR	24
5	Use Cases: CDR for Social/Political Good	26
5.1	Measuring Electoral Fairness During Election Periods in a Voting System	27
5.2	The Use of Social Network Analysis in Call Data Records for Relationship	
	Mapping and its Implications	28
5.3	Socio-economic Wealth Distribution Using Call Data Records	29
5.4	Data Insights on the Dynamics of Peoples' Migration — Before and After	30
5.5	Travel Patterns and implications of Border State Sojourners	31
5.6	Mobile Call Volume Forecasting of New Settlements	32
6 U	se Cases: Commercial Applications	33
6.1	Call Driven Insights Campaign for Mobile Operators	34
6.2	House Prices Analysis and Prediction using CDR	35
6.3	Urban Activity Pattern in CDR Data	36
6.4	Ride Sharing Opportunities using CDR Data	37
6.5	Crowd Estimation at a Social Event using Call Detail Records	38

7 U	se Cases: Governance	39
7.1	Planning for sustainable cities by estimating building occupancy with mobile phones	40
7.2	Intelligent Distribution of Movable Amenities: A Case for Timely	
	Redistribution of Mobile Toilets	41
7.3	Decongesting Markets and establishment of New Markets. Where and When?	42
7.4	Tracking Work Related Movements During A Lockdown	43
8 U	se Cases: Health and Society	45
8.1	Improved Disaster Response Management using Mobile Phone Network Data	
	— Human Mobility Pattern Tracking in Northern Nigeria Using CDR	46
8.2	Mental Health State using Insights from Call Detail Records	47
8.3	Estimating the Extent of Air Pollution using Call Detail Records	48



Introduction

HUMAN MOBILITY DATA FROM ANONYMIZED MOBILE PHONE CALL DATA RECORDS: USE CASES AND APPLICATIONS FOR SOCIAL GOOD

1. Introduction

Cell phone uses has grown dramatically in recent years, penetrating every aspect of modern life. According to Pew_[49], more than 5 billion people worldwide have cell phones, with 2.5 billion being smartphones. Smartphone coverage, however, is imperfect—particularly in the developing world. In emerging economies, it is typical for 40-50% of adults to own a smartphone, with another 20-50% equipped with a mobile phone with fewer features (a "feature phone")_[49]. In emerging countries, on average, 83% of the adult population owns a mobile, 54% of which consists of smartphones. In an advanced economy like the United States, mobile phone ownership reaches 94% overall, with 86% of which consists of smartphones_[49].

The modern cell phone is, in a very real sense, a supercomputer compact enough to fit in your pocket. A handful of decades of innovation have led to the development of devices that perform the functions of a telephone, instant messenger, timekeeping device, calculator, and platform for simple games. In the case of smartphones, these computers additionally serve as internet browsers, personal assistants, delivery devices for multimedia content, high-resolution photo.

The wealth of functionalities built into a modern cell phone comes with a wealth of new data that can produce insight into user behavior at the individual and at the aggregate level. When used wisely and with a focus on service to the user, with respect to the right to privacy and a commitment to using these data for personal and social good, mobile technologies combined with machine learning and data analytics can (1) provide new value to mobile phone users at the individual and societal level, (2) provide new data lakes containing information about aggregate social and geographic trends that can inform governance and policy making, and (3) furnish an analytical and conceptual toolkit that allows insight to be derived from these unstructured data.

1.1 Rationale

This paper provides a compilation of various research and methodological advancement of call detail records (CDR). The research carried out transcends big data available with mobile network operators (MNO) from data occupying computer storage hardware or any storage medium to aggregated and simplified data with rich insights for social good and increased profit margins. Knowing the sensitivity of this data, CDR is required to be adequately annonymised even before data aggregation which further masks any possibility of identification of this dataset. In clear terms, it is an exploration of data rich potentials of emerging

economies (with Nigeria as its case study) using a big data lake of CDR with country's MNOs (MTN, GLO etc) to proffer solutions to socio-economic issues.

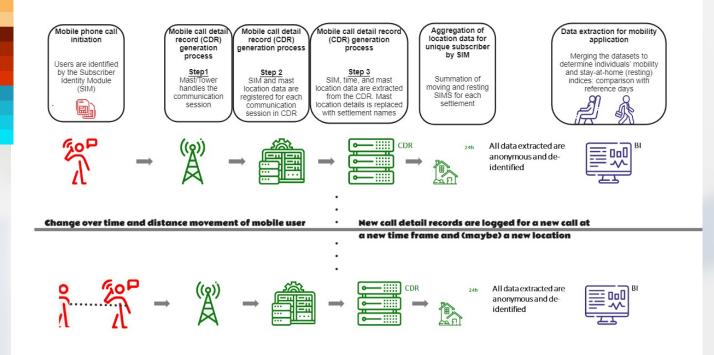
Data Scientists Network (DSN), a non-governmental and research driven organisation, in partnership with Flowminder UK executed the project of exploring various datasets akin to the country's CDR datasets to propose various use cases and have developed protocols for its proposed implementations. The content of the paper gives clear details of various researched areas as outcomes of the project executed.

1.2 Call detail records (CDR) and geospatial leverage explained

Firstly, let's demystify the generation of the CDR dataset. CDR is generated when a subscriber call is initiated (i.e. a phone user makes a call). It is a detailed log/record of a call event of a mobile user by the MNO usually for calculating the monetary value generated. So what makes the simple process relevant? It is the intersection of CDR and cell tower locations dataset that makes it possible to explore human movement data which can be applied to uncover many insights. The pattern of movement of a single subscriber can be deduced when the same subscriber moves to a new location Y position and makes a call. See Figure 1.

The duration of the call is important to MNOs, so it is to data analysts. Also, of more relevance is the centroid location of the caller at the time of making/receiving a call based on the connecting cell tower spatial location for enabling the mobile call. The key data features (or variables) required are — 1) unique subscriber or caller identification, 2) time-stamp and duration of the call or phone activities and, 3) co-ordinates of cell towers for enabling call (or other) events. The coarse spatial information of cell towers is aggregated into names of settlements at various levels of granularity given as — level 0 (country), level 1 (region or states), level 2 (districts or local government area), and level 3 (municipality or street area). These are triangulated to provide further rich insights, which we have documented in this paper.

(A) CDR generation



(B) CDR and geospatial leverage explained



Figure 1: CDR and Geospatial Leverage

1.3 Project Acknowledgements

We appreciation the Bill and Mellinda Gate Foundation for their support for this research work on CDR and geospatial analytics for Nigeria as a case study. We also acknowledge our partner, Flowminder UK, for the success of the project.



2. Roadmap

This report provides a survey of use cases for mobile call record data with a particular orientation towards using these data for social good. Machine learning has a profound and largely untapped potential to transform the way we process information at scale and design systems that automatically produce insight from the enormous amounts of data produced daily in the digital age.

Realizing the transformational impact of machine learning in this space requires an understanding of the technical aspects of how these systems are deployed: (1) the nature of the available data, and (2) the technical tools—algorithms and analytics methodologies—that can be used to build predictive models and extract insight from the data.

Truly integrating these technologies into social usefully systems, however, requires in addition a thorough understanding of the use case landscape, one that brings in all relevant stakeholders to define a problem space, envisions potential solutions at a high level, and adapt the solution concept to the domain requirements as expressed by stakeholders. Ensuring that proposed new systems are precisely fitting the needs of potential users is an absolute must for unlocking social value from data and algorithms, and requires careful and deliberate process mapping associated with each potential use case. In this report, we aim to provide not just an overview of the technical capabilities/requirements associated with each implementation domain, but also

an understanding of how these systems would operate in the real world, and the necessary adaptations of the Al toolkit to fit the needs of an end-user—whether that user be a business, an NGO, a humanitarian relief organization, government, or individual.

2.1 Stakeholders

The use cases explored in this report are aimed at a broad variety of stakeholders with a particular emphasis on applications useful for Africa and Nigeria. Readers are encouraged to jump around to sections within their application domains. Regional and municipal planners might be interested in how CDR can be used to improve transit systems (Section 4.1) or waste management amenities (Sections 3 4.2 and 7.2) using fine-grained population flow maps built from call volumes.

Government officials with a sustainability portfolio might be interested in how CDR can be used to dynamically infer a building's precise electricity needs throughout the day using occupancy imputation models (Section 7.1). Housing developers could hop to Section 6.2, where we show how CDR has been used to predict local housing market conditions using machine learning. Investors might wish to envision new CDR-based services that could provide personalized rideshare recommendations based on a user's mobility records (Section 6.4).

Public health officials in Africa might be interested in investigating lockdown compliance during COVID-19 to improve future epidemic response (Section 7.4). National security experts and NGOs could wish to learn how to improve election security and fairness by tracing anomalous population movements during a national vote (Section 5.1), or how to better provide for refugees fleeing violence and insurgency using population flow modeling (Section 8.1).

In each use case, we map the social, policy, and business landscape as well as the technical tools necessary to unlock the untapped value of CDR.



3. Scoping the Data Landscape

Mobile call records or call detail records (CDR) refer to the array of metadata associated with a phone call or other communication via telephone network infrastructure. Call detail records are produced whenever a mobile provider transfers digital information across their telecommunications system.

CDR records may include fields such as:[24.20]

- 1. The phone number of the originator and the recipient
- 2. The starting and duration of a phone call
- 3. Timestamp for a text message
- 4. Geospatial information associated with the billing account (e.g. home address)
- 5. Coarse geolocation information associated with the call based on coordinates of nearest cell tower

We focus on this set of datapoints because, in contrast to high-resolution Global Positioning System (GPS) data, CDR is truly universal in that it can provide useful geolocalization for nearly all mobile phone users. This is of particular importance in emerging economies, where only about half of mobile users use GPS-enabled smartphones, and also in applications where comprehensive samples are required without inequality-based selection bias caused by the relatively high cost of smartphones.

According to Statista, Nigeria has 40,541 cell towers $_{\tiny{[50]}}$, with an average density of one for every 8.8 square miles. In practice though, it is possible to get much higher per-call spatial resolution since most calls

emanate from large population centers with a greater density of cell towers, making per-user and per-call geolocation much more accurate under optimal conditions.

Population mega centers like Lagos State have 100% geographic cell coverage, including for high-bandwidth 4G telenetworks. See Figure 2-4 According to the International Telecommunications Union, the accuracy of access tower-based geolocation from CDR varies from 50 to 300 metres in dense urban areas.

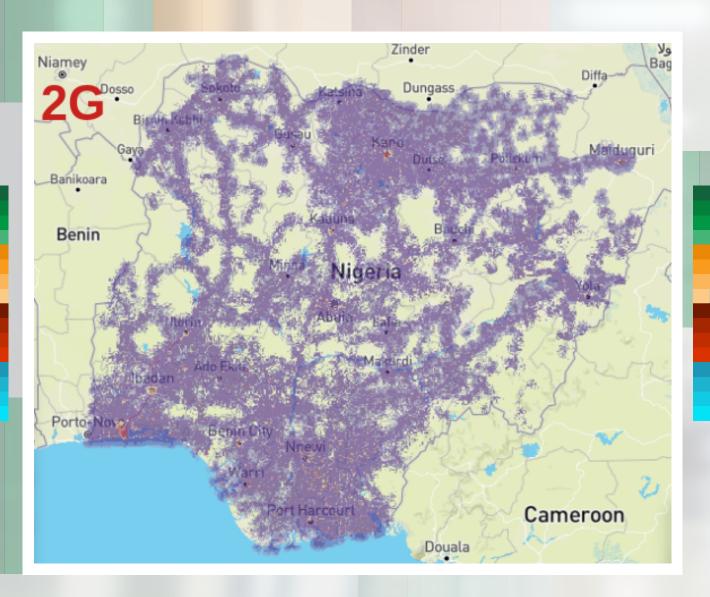


Figure 2: 2G, 3G, and 4G mobile coverage maps for Nigeria.

Data source: Facebook Connectivity Lab and CIESIN (Columbia University) via www.mobilecoveragemaps.com



Figure 3: 2G, 3G, and 4G mobile coverage maps for Nigeria.

Data source: Facebook Connectivity Lab and CIESIN (Columbia University) via www.mobilecoveragemaps.com



Figure 4: 2G, 3G, and 4G mobile coverage maps for Nigeria.

Data source: Facebook Connectivity Lab and CIESIN (Columbia University) via www.mobilecoveragemaps.com

By the same token, CDR can be used to build activity indicators not captured by GPS alone, such as time-varying measures of social engagement (Section 5.2), and direct indicators of user wakefulness and daily

routines that can be useful in applications like tracking aggregate mental health (Section 8.2). CDR thus offers challenges and opportunities unique to this technology.

Challenges	Opportunities
Spatial resolution: Generally on the order of kilometers	 Population coverage: Smartphone + Feature phone coverage is generally well above 80%, including in the developing world
Temporal resolution: Data only available when user places a call or sends a text	Sidestepping inequality-based bias: High cost of smartphones means GPS data only sample wealthier users
Sampling sparsity/bias: Only geolocalized during waking hours/social activity	Spatial resolution: Kilometer level resolution data is sufficient for many aggregate analytics
Interpretation: Requires knowledge of & sensitivity to typical daily routines of mobile users	Temporal resolution: GPS oversamples relative to analytical needs, leading to immense data lakes that can become computationally intractable
	New measures: Sampling from call/text records provides measures of social engagement, wakefulness, mental health

3.1 Privacy

While CDR has found many applications in governance, particularly law enforcement, access to such records appropriately has to go through multiple layers of ethical or judicial review₁₇₇. When handling call record data, it is

of utmost importance to protect the privacy of the user. The required steps differ based on the application. Geolocation data can be extremely useful in the aggregate, where individual user identities are not

needed in order to derive insight from data. For instance, estimating location-based population density at different times of day (Section 6.5), or tracking infrastructure development needs through the analysis of commuter mobility patterns (Section 4.6), are all analyses that can be conducted after extensive anonymization, and removing personally-identifying call record information such as individual phone numbers or

billing address. In order to build trust with mobile users and ensure that their expectations for privacy are met, CDR data lakes intended for aggregate analysis should always undergo such preprocessing. Applications requiring privately held personally identifying information should always obtain affirmative consent prior to collection/use.

3.2 Leveraging AI and Machine Learning for CDR Data

Machine Learning methods are ideal for navigating the immense amount of call record data generated every day by call detail records. Modern machine learning algorithms with billions to hundreds of billions of parameters, and can tractably be trained on billions to trillions of datapoints. The democratization of Al through platforms like Pytorch, Tensorflow, OpenAl and Hugging face have contributed immensely to the availability of state-of-the-art Al methods for the analysis of a broad range of data, enabling analysis- and service-provision tasks at a volume and scale that would be impossible for human workers to deliver.

Models suitable for processing CDR data range from simple volume statistics to build relative geographic density maps (e.g. Section 6.3) to advanced statistical models like Page Rank-like caller-recipient graph, eigen decomposition to predict influencer centrality in social networks (Section 5.2) or neural network-powered predictive algorithms for tasks like forecasting housing prices (Section 6.2). Many of the data-gap problems associated with call records can be overcome with

techniques like trajectory interpolation [33] essential for building fine-grained spatial maps of mobile user movements. Even simple statistical algorithms like linear regression are generally needed to provide a linking function between CDR volume data and predictions of real-world variables like crowd size at an event (Section 6.5), and can work exquisitely well at predicting real-world values from CDR observations as long as ground-truth training data are available for fitting the models.

When delving into the world of call record analysis, stakeholders should be aware of the characteristics of CDR data, as well as the techniques (Models, Algorithms, and Machine Learning Architectures) that can be used to build predictive systems or extract useful insights. Table 2 provides a survey of key terms that come up in call record analytics, and are useful for understanding every step of the pipeline from preprocessing call records to deploying them in predictive forecasting.

Term	Definition
Machine Learning	The discipline of using data combined with algorithms to develop and train models to perform inference on datasets resembling those on which the model was trained. Machine learning typically involves specifying a statistical model of the data with associated input variables or features that are informative for predicting outcomes using the model.
Algorithm	In the context of machine learning, an algorithm refers to a statistical model and the process that is used to obtain parameter estimates that are a function of the input data. For example, when training a neural network, the parameters are the numerical values of weights between units in the network. In order to estimate the parameters that optimize performance on the target task, we would use the training algorithm of stochastic gradient descent, which helps the network produce predictions and incrementally adjusts the network weight parameters so as to maximize the accuracy of those predictions.
Model Selection	The process of choosing a model among an inventory of potential variations on that model. Machine learning distinguishes between parameters—whose values are derived from the training algorithm applied to the data—and hyperparameters—which are manually set by the model designer. In model selection, the architect tests different combinations of hyperparameters and chooses the one that maximizes performance on a set of new data (data on which the model was not trained) in order to avoid overfitting. Model selection also involves testing and choosing which features are genuinely important to model performance.

Term	Definition
Call Data Records (CDR)	Data associated with individual communications across a mobile telephone network—including phone calls and text messages. CDR does not contain content associated with communications, but only the metadata of a call, including time, location, and duration.
Global Positioning System (GPS)	High-resolution geolocation data based on satellite arrays typically maintained by governmental entities. In contrast to CDR, GPS samples at an interval of seconds and with a spatial resolution on the order of meters. However, a large share of mobile phones are not GPS-enabled.

Table 2: Key Terms in Machine Learning and Call Data Record (CDR) analysis.



4 Use Cases: Organizational/ Infrastructural Optimization using CDR

4.1 Optimizing Public Transportation System Using Subscribers Density and Point of Interest

Many cities struggle with building and maintaining public transit systems that meet the needs of the broad array of public transit users [7]. A public transit system needs to simultaneously satisfy the needs of residents, commuters, and visitors who use the system, as well as those who are not regular users of public transit (residents and drivers who experience the positive and negative externalities of public transit traffic). Building smart public transit systems requires a balance between speed and efficiency of transit, directness of route without too many layovers at a "hub", as well as normative goals like preserving residential, commercial, and tourism neighborhoods.

Technology-assisted mapping using Geographic Information Systems (GIS) has been successfully deployed in estimating key parameters for a public transit system in Lagos, Nigeria, including optimal number of round trips, number of vehicles required along a given route, and expected revenue [7]. However, the data available for community mapping in service of building smart transit systems is often limited to

qualitative observations from field surveys and actual usage statistics from existing transit system operations. While usage data can be useful for incrementally adjusting bus routes, it does not by itself provide insight into potential transit endpoints that are not integrated into the existing system. A natural urban point-of-interest that is beyond walking distance of nearby transit stops will lead people to use alternative transit methods like personal vehicles, leaving this site out of the usage-data feedback loop.

In the absence of data about potential usage, the opportunity to extend the transit system to this valuable new site would be missed. Furthermore, while it may be possible to adjust overland bus lines in response to usage statistics, smart transit planning for systems with high associated capital expenditures—such as a subway—simply cannot take a costly "build-and-see" approach.²

"Understanding which areas have a greater volume of calls at a certain period of the day can inform real-time decision-making in road transport management and control."

² https://www.cnn.com/travel/article/chongging-china-metro-station-nowhere/ index.html

CDR data can provide an additional input into the transit optimization process by offering time-varying snapshots of location data for large samples of people, allowing for identification of exactly such opportunities. Such data have been used successfully to estimate inter-urban transit times [40]. This application does not fully exploit the spatial specificity intrinsic to call records [18].

By measuring mobile activity with high spatial precision—particularly in urban areas—CDR mapping can provide insight into potential service locations where the current system has no existing coverage. Time-varying measures of call density can also inform route logistics, answering questions such as how many buses/trains per hour are needed to meet the expected for a given route at a given time of day. And since, for instance, people place calls during the day when they are at work and at night when they are at home, CDRbased maps can also directly provide insight into neighborhood characteristics, informing transit system architects about the locations of residential, commercial, and business districts. This enables smart planning decisions that maximize social harmony. For instance, avoiding late-night overland transit activity in residential areas.

Social Impact. Understanding which areas have a greater volume of calls at a certain period of the day can inform real-time decision-making in road transport management and control, benefiting the commuters, allowing for dynamic expansion and contraction of facilities, reducing pollution in residential zones, enhancing social interactions through greater interconnectedness, and managing traffic flow in the state. Smart urban governance [21] empowered by technology can improve life along all these dimensions.

4.2 Precise Waste Management (WM) Using Call Data Records and Satellite Imagery

"CDR-based growth indicators can also be used for long-term planning of Waste Management capacity, for purposes of infrastructure and WM route build-out."

Where there is life, there is waste. Effective waste management is essential to guaranteeing clean air, water, safe foods, and is also a key part of mitigating the environmental impacts of human activity. Waste management (WM) is a core function of any society, and according to the UN Environmental program, is among "the most complex and cost-intensive public services", but also a key factor in sustainable development.3 Central to WM are both individual-level and systemslevel parameters. An efficient WM system must communicate to individuals their responsibilities (collection times, waste disposal protocols such as where to put WM bins, how to separate refuse from recycling, etc.), and also maintain system-wide efficiency in a vast infrastructure that includes a workforce, a vehicle fleet, sorting hubs, and land-fill locations. Problems associated with waste-related pollution are severe impairments to business activity, tourism, and quality of life in affected urban areas.

Technological innovation in waste management includes new devices like Al-powered smart waste bins that automate item sorting, e-Waste kiosks that incentivize responsible disposal with cash returns to users who deposit their waste⁴. Researchers have also

³ https://news.un.org/en/story/2012/11/424932

⁴ https://www.forbes.com/sites/forbestechcouncil/2022/09/09/role-of-technologies-and-mobile-apps-in-waste-management/?sh=53343b027d26

proposed using the capacities of GPS-enabled mobile phones to promote effective fleet management with WM activity tracking, allowing dispatchers to detect jams/bottlenecks, allowing them to re-route WM workers in real-time 1001. CDR data provide additional data inputs that can be used both dynamically during service provision, and architecturally from the point of view of scoping infrastructure requirements and optimizing WM systems. Identifying residential and business hubs through analysis of call patterns (daytime vs nighttime call volumes) can provide population estimates useful in preventing pollution of human-occupied environments by indicating where land-fills and hubs should be placed. These estimates can also be used to predict the type (home refuse vs. business-quarter trash e.g. from restaurants) and volume of waste at particular grid points in a community map, pinpointing where, how often, and how many WM vehicles are deployed in a given service region. CDR-based indicators of population growth can also be used for long-term planning of WM capacity, indicating the need for additional investment building management infrastructure.

Social Impact. CDR-based mapping can become an essential part of the urban planning toolkit, and in the domain of Waste Management, can be of tremendous use to governments and WM service providers in promoting efficient operation. Its social impact goes well beyond service needs, promoting environmental sustainability as well as public health, wellness, quality of life for a group of stakeholders that includes anyone who interacts with WM daily.

4.3 Digital Inclusion and Mobile Penetration in Nigeria Using CDR data

Expanding access to the digital world is a key humanitarian goal of the 21st century. As hundreds of millions of people were forced indoors during the COVID-19 pandemic, we became aware of the tremendous importance of digital technology in ensuring our continued ability to work, learn, and socialize during times of crisis. Crucially, students with internet access were able to continue their education throughout the pandemic, and even outside of crisis times, ready internet access has become an essential component of an education, allowing students to access a volume of information unthinkable even a decade ago, and providing them the tools needed to become proactive learners and digital citizens. Indeed, research indicates that internet use correlates positively with educational attainment [5], even when controlling for family, demographic, and other effects [42]—a principal reason why the UN Secretary General has identified universal affordable digital access as a key development goal by 2030.5

Access to telecommunication networks is essential for education, business, and quality of life. But in order to expand access and promote digital inclusion, current levels of telecommunications use need to be accurately measured. CDR data can be used to obtain accurate counts of the number of unique subscribers per region, which when compared with official government population estimates, allows for the tabulation of digital penetration in a given geographic region, allowing policymakers to set localized targets for expanding access to and awareness of technology.

"CDR can yield realtime indicators of mobile technology use, in contrast to the lag of government statistics."

Since call records also commonly indicate when users are communicating with the Internet, CDR provides direct estimates of the relative rates of smartphone vs. feature phone usage, capturing a dimension of inequality crucial to securing broad access to information and overcoming the digital divide. Using population-wide estimates of current usage, governmental and non-governmental stakeholders can launch strategic campaigns to expand mobile and Internet usage, which can be rolled out both quickly and accurately.

In addition, CDR can yield real-time indicators of mobile technology use, in contrast to the lag of government statistics and years-long sociological processes, particularly in rural areas with underdeveloped state capacity. The presence of mobile activity can be used to impute correlated development indicators like home access to electricity where direct measures may not be available [34], and can be part of an overall strategy of data analysis serving balanced development goals.

Finally, a key aspect of promoting peace, harmony, and knowledge transfer both intra- and internationally is the amount of communication between communities. Mobile technology has the potential to mitigate rural economic and geographic isolation, but this potential can only be realized if there exist tangible social network that users can access in other regions, providing individual opportunities, economic development, and a broadened pool of social interaction, promoting mental

health. Since call records can be matched and localized at the level of both the caller and the recipient, analysis of CDR data can directly yield measures of community interconnectedness and individual/inter-regional isolation, which policymakers and program designers can use to appropriately target efforts aimed at increasing social interconnectedness in areas of need.

Social Impact. Given the importance of communication networks in producing wealth, promoting education through knowledge transfer, and producing digital citizens, overcoming the digital divide is a key goal for both developed and developing countries. Mobile call data records provide a significant data input into mapping current telecommunications usage and access trends, useful for both long-term planning of tech infrastructure expansion as well as targeted campaigns aimed at boosting telecoms use in select areas.

4.4 Police and Security Alertness using Aggregated Call Data Records

The nation of Nigeria faces a complicated set of security concerns. In recent years, the country has experienced waves of sectarian violence and terrorism that have caused trauma and strife, displacing millions of people from their homes. Organized crime poses a major problem for Nigeria's security and development.

Building a peaceful and prosperous society requires people to feel secure in their lives, and effective policing can be part of how security is achieved. Individual-level call record data are a routine part of criminal investigations in many countries, allowing investigators to gather evidence in the course of court proceedings. Tools that allow law enforcement to quickly shift through tens of thousands of phone records have been in use for decades, and are particularly useful in tracing the activities of organized crime networks [25, 38]. Forensic analysis of a suspect's internet and phone records in temporal and spatial proximity to a crime or terrorist attack can also provide evidence of involvement in the event, and can even help preempt commission of a crime [37]. Knowledge that a criminal's data records can be accessed via subpoena in the course of an investigation can even serve as a deterrent to criminal activity.

"Up-to-date population estimates for a neighborhood or district can assist in allocating policing resources as a function of their service area."

At the aggregate level, CDR can be used to predict moment-by-moment population density levels in an area of interest to law enforcement.

Unusual congregations of mobile users in remote areas could provide indicators of armed terrorist activity, providing an early warning system for condensed criminal activity. Police conducting operations in a certain area can benefit from additional information about the number of people on the ground in the target location, which can be obtained from call/text-volume statistics gleaned from the CDR steam. Up-to-date population estimates for a neighborhood or district can assist in allocating policing resources as a function of their service area.

Social Impact. Achieving personal and national security is a goal for all nations, and Nigeria has struggled with this in the past. By appropriately leveraging individual and aggregate calls records in compliance with judicial procedures and the right to privacy, law enforcement agencies can enhance their activities in fighting terrorism and organized crime, leading to a safer society for all.

4.5 Optimizing Electricity Allocation By the Dynamics of People's Movement

Medium- and longer-term planning of electrical generation capacity requires the ability to dynamically assess population growth and mobility at time scales of months of years. While online usage data are useful from the point of view of moment-to-moment allocation of electrical generation, essential decisions about what electrical infrastructure (power plants, power lines, grid access nodes) must be made years in advance on the basis of limited information about population density and usage trends. Census-based estimates of regional population levels based on census data often lag behind the actual dynamics of population movement and growth, making it difficult to plan for quick and efficient allocation of infrastructure.

"CDR can help policymakers and power systems planners to plan for additional power demand at hourly scales, informing the placement of peaker plants and distribution networks in a dynamic way."

https://theconversation.com/why-nigerias-electricity-grid-collapses-and-how-to-shore-it-up-179705

This is of particular importance in a country like Nigeria, whose national grid has collapsed more than 200 times in the last nine years, leading to widespread blackouts and a tremendous human and economic cost.

Particularly at a time when an aged electricity infrastructure requires significant expansion and upgrade, economic planning for new capacity requires a dynamic understanding of population trends in electrical service areas. Given the near-ubiquity of mobile phones among adults in Nigeria, CDR data can yield accurate population estimates useful to power infrastructure architects.

The additional temporal resolution of CDR in comparison to census-based estimates can also be of tremendous use in predicting, for example, temporary population growth in tourist centers requiring additional generation/distribution capacity during the summer months. As well as predicting the locations of residential /commercial zones in a way that can inform the need to produce additional peak-generation capacity to serve needs that change hourly as people commute between work and home.

Social Impact. Efficient electrical allocation is of key importance to providing for the needs of a national population, and countries like Nigeria have recently struggled with establishing a smart electrical grid. CDR can help policymakers and power systems planners to plan for additional power demand at hourly scales, informing the placement of peaker plants and distribution networks in a dynamic way. Seasonal and hour-scale population estimates derived from CDR can enable electrical generation to follow users from city-to-city, or from home-to-work.

4.6 Traffic Analysis and Prediction using CDR

While individual-level geolocation sampling frequency in CDR depends heavily on the user's communication patterns (how often do they place a call or text message?), particularly active users can provide extensive samples of individual mobility trends.

In dense urban areas, user location can be pinpointed with accuracy between tens and hundreds of meters, and CDR data have been successfully used to establish user-level trajectories with fine spatial resolution using machine learning based point-sequence estimation with spatial interpolation, allowing analysts to build maps of exemplar trajectories on the basis of high-frequency telecoms network users.

Trajectory maps of this kind can estimate not only location, but also identify stay-points where a user spends a great deal of time, as well as predicting mode of transportation (walking, car, bus) on the basis of the magnitude of user movements across time (rate-of-travel estimation).

"CDR data can be used to provide individual-level services like traffic map provision and best route recommendation that give travelers additional power to reduce their travel times."

These measures are useful for both individual-level and aggregate-level tasks. At the individual level, momentby-moment estimates of predicted traffic volume, specific to a geographic area and to a transit modality (pedestrian vs. vehicular congestion) can be used to provide maps of predicted traffic levels in an area. A user might query, via an app, what is the predicted amount of traffic along a given transit route, using CDR-based estimates to guide their decision of whether to walk to a location, take the metro, or their car. Depending on the developer's amount of access to real-time CDR data, traffic mapping could be based on direct sampling from the current CDR-stream, or using predictive modeling based on prior snapshots of CDR data, using variables like time-of-day, time-of-year, and weather (pedestrianfriendly or not) to predict traffic patterns moment-bymoment using machine learning tools cross-validated on a large sample.

It is worth highlighting the contrast between CDR- and GPS-based traffic estimates of individual-level location. While apps like Google/Apple Maps and Wayz offer estimates of traffic congestion based on high-resolution GPS data, their estimates depend on a sufficient base of users who (a) have GPS-enabled smartphones and (b) use the app. This raises issues in countries like Nigeria, where only 39% of the population possesses a smartphone⁸, making CDR-based traffic modeling potentially more accurate in the developing world, particularly in the context of high inequality and limited smartphone ownership.

At the aggregate level, CDR can facilitate analysis of traffic trends that may be of use for infrastructure planning and mapping: where to build highways or additional highways lanes, where to build/extend public

transit lines (see Section 4.1), or predicting high-level population trends requiring brand new transit infrastructure. Windowed snapshots of current traffic activity can assist in planning particular infrastructure projects, by indicating dynamic levels of route usage along major thoroughfares, allowing construction crews to plan building projects and road closures accordingly in a way that is sensitive to the needs of road users and also the locations of residential areas.

Social Impact. Travel time is a major mental health hazards for commuters [39,32], and is a particular problem in places like Lagos, Nigeria, where traffic jams can routinely add hours to a commute [6]. CDR data can be used to provide individual-level services like traffic map provision and best-route recommendation that give travelers additional power to reduce their travel times.

Government infrastructure planners will also benefit from the use of spatially and temporally refined traffic maps, benefiting public health and economic efficiency.



5 Use Cases: CDR for Social/Political Good

5.1 Measuring Electoral Fairness During Election Periods in a Voting System

Many countries suffer from poorly organized and monitored electoral systems with low transparency. This leads to electoral corruption and loss of faith in national political systems, with major knock-on effects on democracy, governance, public wellness, and economy.

A deficit of democratic fairness and electoral transparency is particularly problematic among lowand middle-income countries. Mechanisms of electoral fraud in Nigeria historically have included manipulation of the electorate, gerrymandering, vote-counting fraud embedded in voting machines, demographic manipulation including manipulation of immigration laws, disenfranchisement through selective purging of voter rolls, and voter intimidation/electioneering and polling sites. Local dynamics of electoral corruption can lead to the governing parties to mobilize groups of voters to cast ballots outside of their home jurisdictions in exchange for cash.

Authorities have been known to move populations to different voting locations by temporarily assigning land deeds or moving their residence through public housing low. While many aspects of monitoring elections for freedom, transparency, and accurate counting often require extensive on-the-ground presence by impartial election observers, electoral anomalies can often be

detected through aggregate geographical data. Mobility before, during and after an election can provide signs of electoral manipulation through vote-buying and voter geographic reshuffling in the electioneering process, particularly when voters are moving across state boundaries. For instance, a sudden surge of CDR activity in a jurisdiction can indicate that the governing party has manipulated voters to change jurisdiction in order to win particular seats. A CDR-based election fairness tracker would trace mobility at least a month before, during and after the election process in a ballot voting system. CDR-based population estimates can also be used to detect unexpectedly low ballot returns in a region, indicating that ballots may have been "misplaced" through corruption.

Social Impact. Trust in elections is a key feature of successful nation building. By providing additional tools for national and international election monitors and security agencies to detect anomalies in electoral activity, CDR analytics can help identify political corruption, build stronger governing institutions, and ensure trust in the integrity of current voting systems.

"Mobility before, during and after an election can provide signs of electoral manipulation through vote-buying and voter geographic reshuffling in the electioneering process, particularly when voters are moving across state boundaries."

² https://www.cnn.com/travel/article/chongging-china-metro-station-nowhere/ index.html

5.2 The Use of Social Network Analysis in Call Data Records for Relationship Mapping and its Implications

Social networks are an essential part of a good society. Atomized individuals are less happy and less healthy, and find themselves unable to access opportunities available to those with a greater number of connections to others. Dynamics of social networks are important not just at the individual level, where they impact personal health and economic prospects, but also at the aggregate level.

Communities with high density of connections to neighboring communities are able to build social and economic links, fostering unity and enhancing local economic activity. CDR data provide natural metrics of inter-regional connectedness through bilateral localization. When records are available for both parties to a call or SMS, it is possible to map which regions are communicating with which other regions at the aggregate level through Social Network Analysis_{IRI}.

Each subscriber represents a node in a graph, each node associated with static features like a home address and also dynamic features such as location when a call/message (graph edge) is placed. Nodes can be clustered together on the basis of geographic coordinates, yielding "super-nodes" whose level of interconnectedness to other nodes can be aggregated

"CDR data provide natural metrics of inter-regional connectedness through bilateral localization."

by counting the number of edges to other communities. Graph-based affinity levels between individual communities can be used to analyze community-level isolation and interconnectedness, which can be essential for studying the effect of social networks on health and economic indicators. Beyond providing new opportunities for research, such analyses can help societies overcome social and cultural barriers like rural-urban divides, and promote interchange between different linguistic, cultural, and demographic groups who might engage in conflict in the absence of opportunities to build cross-group empathy.

Metrics of this sort have also been used in commercial applications, for instance to identify "influencers" on the basis of network centrality measures (number of incoming and outgoing edges) so as to better target communication campaigns aimed at reaching large numbers of people. Graph theory-based methods like eigenvector centrality measures (8)—which are the basis of the Page Rank algorithm—can go beyond simple edge-counts to estimate the probability of reaching any given node from another, taking account of the fact that centrality within a clique may be less valuable than centrality with surrogate links to the rest of the network from the point of view of propagation. Beyond commercial applications, such metrics can be useful for maximizing the reach of public communications, whether in a business or public interest campaign, or when aiming to reach people during a humanitarian response (see Section 5.2).

Social Impact. In aiming to efficiently reach the maximum number of people, businesses and NGOs, governments would all benefit from the ability to find the most active members of communication networks.

empowering them to spread messages quickly throughout a community. Researchers can use CDR data to answer questions about the impact of social interconnections on key social, health, and economic indicators, which organizations can then use to track and alter trajectories of regional isolation that lead to diminished opportunities and sociopolitical strife.

5.3 Socio-economic Wealth Distribution Using Call Detail Records

A key virtue of CDR as opposed to GPS-based analysis is that it removes a significant source of sampling bias: the fact that GPS-enabled devices (smartphones) are much more expensive than basic mobile devices. When aiming to accurately estimate population-level statistics of various kinds (population counts, mobility measures, as well as more sophisticated graph-based metrics with geo-tags), it is often more useful to get larger and broader, cross-sectional samples of the population than it is to have comprehensive samples of a particular individual's location at every moment of the day. In sub-Saharan Africa, a majority own a mobile phone of some kind, ranging from 91% in South Africa to 75% in Tanzania.

However, smartphone ownership is much lower (51% in SA vs. 13% in Tanzania),° intersecting with both intraand cross-national inequality.

"CDR can fill data gaps in national records, allowing for geographically precise identification of high-inequality areas."

In addition to helping overcome sampling bias for as variety of applications, call detail records can also provide indirect measures of inequality based on subscriber characteristics and features like internet usage. Smartphone users will sign up for internetenabled plans while those feature phones will only make use of call/text features. The amount of internet usage (e.g. numbers of MB or GB of data downloaded) can provide further specificity to these proxy indicators, while also helping measure the intangible human capital that comes from participation digital information networks. Researchers have successfully used mobile user airtime credit expenditures to build fine-grained geographic maps of spending on this core service across regions, as well as measuring the level of inequality within compact regions by comparing spending levels of neighboring individuals 1901.

"A key virtue of CDR as opposed to GPS based analysis is that it removes a significant source of sampling bias."

Indicators of this sort can sidestep the limitations of selfreported income levels from national surveys, and fill gaps in income reporting in nations with substantial underground economies, where sources of income and cash flows may not be reported to the government.

Triangulating CDR-based indicators with official estimates can, indeed, help detect informal economic activity in regions where there is a discrepancy between officially reported cash flows and aggregate use of telecommunication networks.

Social Impact. Socioeconomic distribution refers to the dispersal of the country's GDP and national wealth among its population. It is a major aim of countries to eliminate poverty and tamp down on economic inequality, and many official measures are out of sync with realities like the presence of informal labor markets, particularly in rural areas with limited state capacity. CDR can fill data gaps in national records, allowing for geographically precise identification of high-inequality areas and empowering governments and civil society to craft and target effective antipoverty programs.

"CDR-based migration trajectory mapping in the wake of national disasters can also be used to develop predictive models of population

flows in the age of global warming."

the UN, more than 17 million people were displaced by disasters in 2018 alone of, and by 2050 there may be as

many as 1.2 billion climate refugees" as a result of

weather disturbance, natural disasters, famine and

drought emanating from global warming.

5.4 Data Insights on the Dynamics of Peoples' Migration — Before and After

Internal and cross-national migration can be difficult to track. Immigrants do not always pass through official border locations, and even developed countries struggle with varying estimates of the size of undocumented immigrant populations due to the difficulty of measuring undocumented transnational travel. Just as an influx of immigration can cause problems like misaligned provision of government services based on undercounts of the local population, migration outflows lead to depleted workforces and tax bases. Accurate estimates migration flows are essential for business and national governance.

Natural disasters, ethnic conflicts, and pull factors like higher wages, better employment opportunities, and the possibility of a higher standard of living are perennial drivers of transnational population flows, and it is expected that migration will increase substantially in the coming decades due to climate change. According to The low cost of a feature phone places mobile access within the hands of many migrants, and both new subscription and dynamic usage data can be used to estimate the inflow of immigrants by comparing usage volumes at different times. The development of new immigrant communities can be tracked at the regional level using CDR geolocation. Data sharing agreements among neighboring countries would allow for building maps of aggregated individual migration trajectories, allowing for precise identification of the sources and targets of migration. CDR-based migration trajectory mapping in the wake of national disasters can also be used to develop predictive models of population flows in the age of global warming—for instance, yielding computational systems that can forecast the magnitude and direction of new migration likely to occur in the wake of a famine, drought, or hurricane. Real-time data of this kind can inform agencies like the National Emergency Management Agency (NEMA) as to where resources must be placed in order to respond to current and future emergencies.

Social Impact. Immigrants can provide tremendous social, cultural, and economic wealth to the lands where they settle. However, in order for society to properly provide for immigrant communities and also reap the collective benefits of their presence in the nation, immigrant communities need to be accurately

5.5 Travel Patterns and implications of Border State Sojourners

Urban centers tend to be magnets for business activity—whether serving as a central location for office buildings, or as commercial centers where consumers from surrounding areas aggregate to purchase goods and services only available in the city. While many cities experience daily ebb and flow of commuters, this dynamic is particularly acute in African megacities like Lagos — the second most populous in the continent — where housing prices are beyond the means of many regional workers. This has led to the phenomenon of border state sojourners — commuters who traffic in to Lagos, many from other nations.

According to the Lagos state governor, as many as 123,000 border state sojourners may migrate to Lagos every day, many of whom come to the city to work and return to their border homes at the end of the day. Estimates of this population flow are often based on data like mass transit usage statistics (incoming buses and trains), which do not reflect the full reality of daily migration into Lagos. A poor understanding of border state dynamics in Lagos leads to strained public service provision (e.g. public health, transit, waste management, and public safety/policing systems), with municipal

systems that are not carefully tuned to the needs of city residents and the sojourner workforce.

CDR data can be used to track mobility pattern of people moving from border towns into the city, providing more reliable estimates than narrow mass transit usage statistics or costly land surveys. CDR data have been successfully used to impute geographically refined mobility patterns and identify daily stay locations like work and home, with aggregate measures in excellent agreement with survey data but at much lower cost.

"CDR data can be used to track mobility pattern of people moving from border towns into the city, providing more reliable estimates than narrow mass transit usage statistics."

Daily and nightly call volumes can be analyzed to precisely localize sojourner communities and ensure that they are provided with necessary infrastructure and government services, in light of their essential role in the social and economic vitality of Lagos state. They might also be used to assess latent demand for resources like housing closer to the heart of the city, allowing for long-term planning of expanded residential facilities for current sojourners commuting from border towns, drawing day-migrants to become permanent residents and allowing the state IRS to tap into a new base of taxpayers who benefit from greater proximity to their place of work.

Social Impact. Border state sojourners are part of the lifeblood of many mega cities, but are often not sufficiently provided for, and represent a potential tax

and residency base that is under-recognized by national and municipal statistics, leading to social and transit systems that are not optimized for the needs of such workers. Gaining and understanding of the dynamics of border state migration flows will be useful government policy makers, urban planners, and the IRS, among other stakeholders.

5.6 Mobile Call Volume Forecasting of New Settlements

"New call volume in an area can serve as an indicator of regional population growth."

Particularly in rural areas in an emerging country like Nigeria, new settlements can pop up quickly and informally, often "off-the-grid". The growth of such settling activity can be difficult to measure in regions with limited state capacity, and detecting it has been a particular problem in Nigeria, where geographical records are often inaccurately and infrequently mapped, and poorly stored. Data from ground surveys and satellite imaging are costly to gather and process, and often unavailable even to national governments, leading to large geographic regions that are incompletely mapped with respect to human activity.

Mobile phone penetration in Nigeria is high, with more than 80% of adults in possession of a feature phone. New call volume in an area can serve as an indicator of regional population growth, whether from subscribers setting up residences in a new settlement, or from visitors to new townships that have not yet been linked in to electrical or water systems. Beyond detecting new communities, geographic data enrichment using CDR can also provide insight into the nature of a new settlement—for instance, by identifying a new "dormitory town" of commuters developing in close proximity to an economic hub. An increasing volume of nighttime calls from an unmapped community can signal the need to expand public services, road/railways, and food distribution to a new community, as well as mapping new business opportunities.

Social Impact. New settlements can often spring up outside of the view of conventional population measurement systems, leading to new communities that are disconnected from infrastructure, are insufficiently provided for in resource allocation, and can easily become socially marginalized. Governments and NGOs can use CDR data to quickly identify new settlements in order to identify new areas of need—whether for grid integration or targeted programs like routine immunization. Real estate developers can use CDR to track areas of unrecognized population growth, finding opportunities to provide new housing.

"Real estate developers can use CDR to track areas of unrecognized population growth, finding opportunities to provide new housing."



6 Use Cases: Commercial Applications

6.1 Call Driven Insights Campaign for Mobile Operators

Often, Mobile Network Operators (MNOs) do not make full use of the data at their fingertips. MNOs are the source of call records, but they can also be the beneficiaries of these data, which provide insight into subscriber locations as well as usage patterns across their service areas. Data-driven insights from mobile call records can also unlock tremendous value for society, indicating areas where there is a mismatch between existing telecoms infrastructure and potential telecoms demand, which can be used to expand participation in the digital economy and boost national GDP. Promoting the growth and diversity of MNOs can increase connectivity, introduce economies of scale, and provide consumers with more options for services and providers.

"CDR data analytics can also be used to improve services for existing subscribers and promote retention through predictive modeling."

Coarse-level geographic data about subscription and usage levels in a geographic area can be compared with census figures to estimate the number of people who are not yet using mobile services. Precise geolocation of subscription deficits is useful for a variety of purposes,

from planning expansion of grid capacity and cell tower density in response to population growth, to tailoring outreach campaigns to communities of potential subscribers.

This is particularly important in a country like Nigeria, which has 37 federated jurisdictions in which more than 500 native languages are spoken. Geographically pinpointing potential demand allows an MNO to adapt media campaigns to local language and culture, as well as recruit native speakers for on-the-ground kiosks and business campaigns. CDR data analytics can also be used to improve services for existing subscribers and promote retention through predictive modeling. For example, Verbeke et. al. [52] use CDR-derived features to predict customer churn—the propensity for a customer to drop their current service provider. A major reason for investing in this area is that acquiring new clients can cost 5 times more than retention (na), and the relative cost of retention can be further driven down by identifying potential churn and predicting which subscribers are most likely to be retained by MNO interventions [52]. Machine learning classifiers like decision tree ensembles and support vector machines, combined with rigorous feature and model-selection protocols have shown that billing- and volume-based measures (number of calls/texts) as well as social network data that can be mapped using caller-recipient graphs [22] predict customer churn, allowing MNOs to appropriately target retention efforts.

² https://www.cnn.com/travel/article/chongging-china-metro-station-nowhere/ index.html

Social Impact. Expanding participation in the digital world and growing national economies are priorities for governments and societies around the world. By applying machine learning models and evaluation procedures to CDR data, telecoms companies can retain and grow their subscriber bases, ensuring that all benefit from global connectivity and the digital revolution.

6.2 House Prices Analysis and Prediction using CDR

Housing prices are affected by a variety of factors, including local amenities like schools and parks, proximity to employment and transit hubs, as well as the dynamics of population growth and mobility. Variables like the amount of local economic activity, the presence of nearby employers, or the availability of mass transit options that allow workers to commute in all impact the price of housing in a given area, which affects the quality of life for rent-paying residents and homebuyers as well as the capacity for a community to grow in response to new housing builds and economic opportunities.

Predicting local housing market conditions is important for government (e.g. urban planning agencies) as well as private actors in industries like banking and insurance in order to calibrate local services and prices. CDR can provide an effective informational input into housing market modeling. High-level indicators like number of subscribers in a given area can provide a dynamic picture of local population counts with fine spatial resolution and much higher temporal frequency than a census. Variable rates of mobile activity across the day can also be used to localize residential vs. employment/commercial zones, since daytime calls are

more likely to be placed at work and nighttime calls tend to occur at home. Research has shown that individual-level social/behaviour patterns can be established using call records. CDR paired with machine learning has been successfully used to predict which migrants will move into poorer neighborhoods as a function of the amount of money in their mobile accounts as well as geosocial mapping of their links to people in a given neighborhood [41], from which aggregate inflow can also be predicted.

"CDR can provide an effective informational input into housing market modeling."

Neural networks using CDR-derived features have also been used to directly predict housing prices [47]. Measures of work-to-home distance can be scraped from time-windowed call-location records, and other more sophisticated measures can be derived from CDR geolocation. Entropy refers to the "flatness" of a probability distribution. A mobile network user who places most of their calls at a single cell tower has low entropy, while one who whose calls are equiprobably serviced by a number of local towers has high entropy. Users' movement between multiple cell towers provides an effective proxy for the presence of amenities like restaurants and parks that would draw dwellers out of the home and into the local neighborhood. Such features have been shown to significantly predict mean housing prices for out-of-sample regions.

Social Impact. Planning housing development and providing banking/insurance services to a given geographic region benefits from the ability to predict the dynamics of local housing markets. Housing price modeling can also be used to provide renter/buyer-oriented services by forecasting local housing price

increases that affect budgeting, and inform government regulation of housing prices to constrain rent increases that could displace residential communities.

6.3 Urban Activity Pattern in CDR Data

"CDR can be used to distinguish residential from business/ commercial zones based on when calls are made (during the day or at night), and can even be used to identify, in a fine-grained way, commercial or recreational areas based on weekend activity."

Urban telecommunication networks feature dense cell tower placement, allowing user activity to be localized with good spatial precision on the order of hundreds of meters to several kilometers [54]. It is possible to map not only user locations at particular time stamps, but also user trajectories across time by tracing the set of cell towers accessed across the course of a day.

Data extracted from CDR require careful interpretation due to the fact that locations/trajectories are only sampled when a device makes use of the network, but in practice this frequency can be quite high, since network access involves the full gamut of mobile activities—placing a call, sending a text message, or accessing the internet with a smartphone. The sampling frequency can in fact be on the order of minutes for highly active users. Probabilistic statistical models like Relational Markov Networks can be used to build smooth user trajectories from time series of user activity at multiple access points,

and also aggregated to detect points of interest as well as thoroughfares representing areas of major pedestrian or transit activity [53].

CDR-assisted urban activity mapping can be of tremendous use to city planners and neighborhood /business associations aiming to understand their locality and how best to allocate services to it. CDR can be used to distinguish residential from business /commercial zones based on when calls are made (during the day or at night), and can even be used to identify, in a fine-grained way, commercial or recreational areas based on weekend activity. These data can be used to draw businesses to open up shop in an area by providing quantitative proof of a high density of relevant activity (e.g. demonstrating high weekend traffic), and can inform urban planning efforts to build new water, electrical, or transit infrastructure in areas showing high levels of spontaneous traffic (see Section 4.1). Government land-use boards can also use such data to appropriately zone areas for commercial or residential development.

Social Impact. To build the smart cities of the future, it is essential that planning and governance activities to dynamically adapt to the way residents, businesses, and consumers make use of urban space. Understanding activity patterns of city dwellers can help in transportation planning and management as well as the provision of key social amenities in strategic locations, allowing localities to build out key infrastructure so as to maximize quality of life for residence and expand business activity based on manifest local need.

6.4 Ride Sharing Opportunities using CDR Data

The global ride sharing industry represents more than \$200 billion USD in revenue per year, and is expected to grow by another \$100 billion in the next half-decade¹³. In Nigeria, ride sharing sales have grown by 17.9% in 2022 alone. While ride share navigation and rider localization requires spatial resolution only available from GPS, ride share companies' internal analytics suffer from the fact that smartphone use is limited in the developing world, meaning that modeling factors like latent demand and actual on-the-road traffic is limited to app users with GPS-equipped phones—a small sample of the overall population.

The benefits of ride-sharing include providing opportunities for work, as well as reducing the number of cars on the road, thus reducing traffic and CO² emissions, promoting environmental sustainability. However, the market for ride-sharing is cornered by a small number of large companies, who often pursue exploitative pricing models that use variables like the amount of battery on a user's phone to predict if they will be willing to pay more for a service. I leading many users to seek alternative ways of coordinating a supply of drivers with a demand of ride-seekers.

"CDR can benefit both individuals and society by matching up individuals in the same locality and proposing joint trips through the analysis of individual transit trajectories." New ride-sharing informational services based on CDR can benefit both individuals and society by matching up individuals in the same locality and proposing joint trips through the analysis of individual transit trajectories. Decentralized networks of app users could contribute call record data allowing for the automated analysis of their daily routines, allowing algorithms to match up users with similar starting and end points to link up their trajectories.

While existing peer-to-peer rideshare systems like BlaBlaCar, UberPop, or Flinc allow riders to find each other online, they do not yet automate the process of identifying common trajectories without explicitly specifying user needs [15]. A system with built-in common-trajectory identification could automatically propose ride-share link-ups by identifying work, home, and leisure locations to pair up users with similar needs. Analysis of the user location stream from dynamically collected CDR could also enable online traffic modeling, telling users when their usual route is likely to be blocked or congested and where public transit might provide a better alternative.

Social Impact. Call records provide extensive samples of a user's location throughout the day, and when combined with data-driven behavioral analysis, can be used to design systems that identify shared opportunities for socializing transportation resources like personal cars. Because of the externalities of traffic—such as collectively longer transit times as well as pollution—getting cars off the road is a major goal of governments and societies. Building systems that leverage data to share resources will be an essential part of the cities of the future.

¹³ https://www.statista.com/outlook/mmo/shared-mobility/shared-rides/ ride-hailing-taxi/worldwide#revenue

¹⁴ https://www.forbes.com/sites/nicolemartin1/2019/03/30/uber-charges-more-if-they-think-youre-willing-to-pay-more/?sh=62217b9a7365

6.5 Crowd Estimation at a Social Event using Call Detail Records

Crowd estimation can be difficult and costly, and yet is essential for tracking participation in large social events. Accurately counting the number of attendees at repeated events like festivals, concerts, or football games can be useful for event organizers, marketers, as well as governments that provide traffic and security/policing services for large gatherings. Using CDR geolocation for crowd size estimation requires contending with several limitations of call records. First, not all event participants will access their mobile network during the course of an event, which can lead to undercount. On the other hand, not all mobile users who access a cell tower near an event will have attended the event, which tends towards overcount.

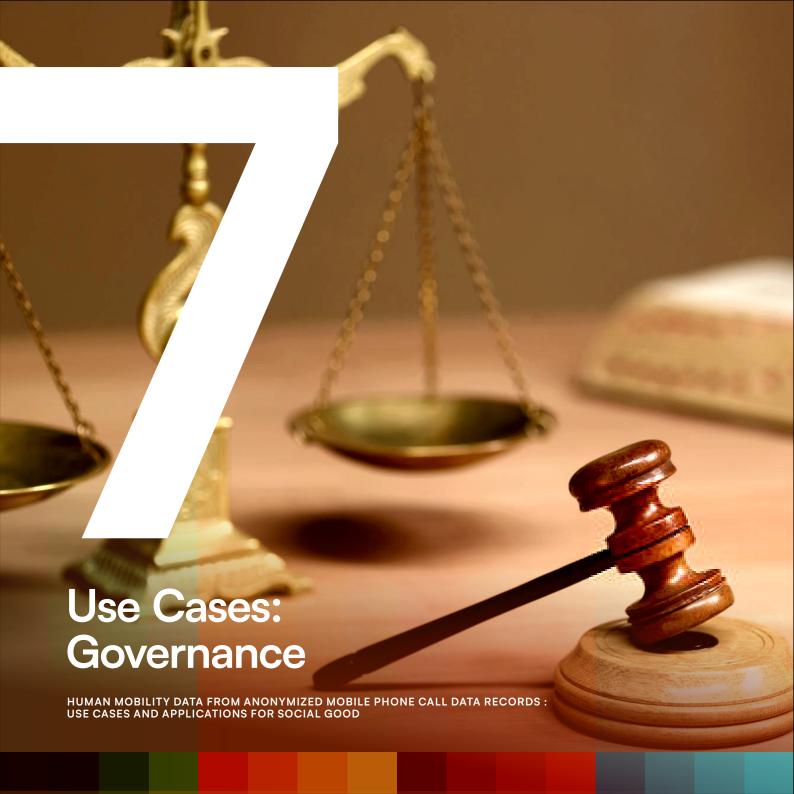
Researchers have dealt with these limitations by using statistical normalization techniques like linear regression to predict crowd size for new events on the basis of events for which a ground-truth sample is available, finding that local-tower call volumes are correlated with event attendance with coefficients upwards of .99, making CDR an almost perfect predictor of the ground truth counts_[5]].

"CDR-based crowd counting correlates extremely well with expensive ground-counts, unlocking a cost-saving application of mobile technology."

Simple attendance counts are useful for providing retrospective analysis of mass social events, and could be sold by MNOs to event organizers like conventions or sports events to help them tabulate interest in past events. The well-validated nature of CDR based crowd-counting, however, allows us to go well beyond retrospective analysis and into predictive forecasting. Hand-counting crowds requires time and personnel, and for this reason many organizers of free-to-view events do not precisely track attendance.

This leads to limitations in the size of datasets available for training algorithms that could predict an event's attendance on the basis of its characteristics (e.g. location, sponsors, the type of event, marketing budget, etc.). CDR can be used to construct observations of event participation for a broad range of events, allowing modelers to train algorithms that predict how popular an event would be, even varying event parameters so as to design the optimal event.

Social Impact. CDR-based crowd counting correlates extremely well with expensive ground-counts, unlocking a cost-saving application of mobile technology for gathering this key retrospective indicator of event success. Predictive modeling can be useful not just for event organizers, but for governments and other groups providing wrap-around services for public events, as well as calibrating access for small businesses (e.g. food carts, tents, and kiosks) to expected attendance.



7 Use Cases: Governance

7.1 Planning for sustainable cities by estimating building occupancy with mobile phones

Building occupancy is inherently difficult to measure. Government statistics are frequently based on small subsamples of residences to yield estimates at the level of a neighborhood or district, so that official counts at the resolution of individual buildings are absent. In many nations and localities, housing occupancy is not officially tracked by government agencies, with informal leasing arrangements leading to even residential occupancy (the occupancy status of a house or apartment) not being systematically tracked. Dynamic occupancy—the number of people in a building as it fluctuates throughout the day—is even more difficult to measure. Innovative research has shown that call data records can be used to accurately predict building occupancy for entire urban populations at scale.

"Call data records can be used to accurately predict building occupancy for entire urban populations at scale."

In a Nature article, Barbour et. al. [11] lay out a framework—TimeGeo—that was used to impute timevarying occupancy at the level of individual buildings using anonymized call records for 1.92 million mobile users in the city of Boston. To overcome the limited

spatial resolution of tower-based localization, TimeGeo uses a probabilistic model of building occupancy that is a combination of observed "stay" points (cell towers where a user spends an extended amount of time) and building-level features like nominal capacity, function (residential/office space), and square footage to probabilistically allocate users within a tower area (TA) to the buildings in that area in weighted proportion to their capacity metrics.

Additional modeling of place-based pull factors like neighborhood popularity are also used to predict effects like "rich-get-richer" dynamics in which, for example, opening a popular restaurant in a district leads to shops and bars opening immediately around it, leading to the development of entertainment districts.

Many modern building functions—such as heating, cooling, and lighting—are or can be handled automatically by smart devices that adapt power draw to expected occupancy throughout the day, leading to energy efficiency savings. Systems like TimeGeo can be used to estimate energy needs which vary with occupancy, allowing for building-level and grid level optimizations that promote lower levels of energy use across the energy system.

Furthermore, accurate predictive energy use modeling is essential in the course of building out new energy infrastructure, and is particularly important when dealing with intermittent energy sources like wind and solar that will serve as the basis for a sustainable world.

Social Impact. Predicting peak occupancy/peak energy use periods is essential to estimating the amount of energy infrastructure needed to provide abundant renewable energy that meets the needs of all users at all times. CDR can thus provide a key input into long-term planning for grid capacity. Going further, dynamic forecasting of energy requirements down to the building level is an essential feature of a future smart grid that can direct energy collected across a distributed network of solar panels, wind turbines, or hydroelectric dams to end-users who can make immediate use of it.

7.2 Intelligent Distribution of Movable Amenities: A Case for Timely Redistribution of Mobile Toilets

Much of the developing world struggles with the need to provide basic amenities for fast-growing populations, including the production of human biological waste. 54% of Nigerians do not have ready access to a toilet, and as many as 46 million Nigerians routinely defecate in the open, according to UNICEF [2]. Open defecation poses major public health and environmental risks. Diarrheal disease like cholera and dysentery—which can be caused by water and soil contamination from defecation in fields, beaches, or bodies of water—are the second largest cause of death for children worldwide, leading to 760,000 deaths per year globally [12]. Hence, the UN has identified universal access to water and sanitation as a top sustainable development goal for the current decade [15].

Some states have outlawed open defecation [4]. However, an approach that criminalizes open defecation, in the absence of providing adequate sanitation facilities for Nigeria's population, would likely only aggravate dynamics of inequality and poverty. As part of tackling the public health crisis, governments have begun exploring the possibility of providing public sanitation amenities, such as mobile toilets, that would tamp down on public defecation while also allowing Nigerians to meet their biological needs.

"CDR-based dynamic population estimates can be used to provide counts of populations within compact geographic regions."

CDR-based dynamic population estimates can be used to provide counts of populations within compact geographic regions, allowing public services officials to set up mobile toilets in locations that are dense in human settlement of traffic flow and poor in sanitation infrastructure. Temporal mapping of human density in service areas can be used to establish cleaning and maintenance schedules by estimating usage volume for mobile toilet facilities—schedules that could vary dynamically by inspecting the CDR volume stream in a designated location.

Social Impact. Open defecation is a major public health risk and intrusion on quality of life, and reducing its prevalence is a major development goal for the world and for the nation of Nigeria. Current efforts are hindered by a lack of sanitation resources as well as a deficit of data as to the optimal placement of sanitary facilities. CDR data can be of use to governments and mobile toilet operators in solving this problem, reducing the amount of roadside defecation/urination and improving life for the affected communities.

¹⁵ https://www.un.org/sustainabledevelopment/water-and-sanitation/

7.3 Decongesting Markets and establishment of New Markets. Where and When?

Open-air markets are a core feature of African culture. There are hundreds of sites across Nigeria where small vendors set up carts and kiosks to sell goods including locally-produced foods and textiles, electronics, and many other product categories. Nigeria is home to the Onitsha Main Market, one of West Africa's largest, and which serves as a cultural and economic center for the city. Open-air markets are an essential part of Africa's food distribution system. In many Nigerian cities, openair markets and informal economies are the main means by which people secure food [30]. Beyond just serving as a point-of-sale for farmers, the Boston Consulting Group estimates that as much as 97% of Nigeria's retail sector—across product categories—resides in the country's 600,000 small traditional retailers operating small kiosks and open-air stands [36,26].

"A market with routinely high density of CDR activity could be expanded to new streets."

Markets can pop up an grow spontaneously, and as part of the informal economy, can be difficult to track. CDR can provide real-time estimates of crowd size with better than kilometer-scale resolution in urban areas, making them an ideal data source for estimating crowd sizes in tightly localized areas (see Section 6.5). Call/text volume at a particular time of day correlates extremely well with the number of people in a geographic area [51], and when combined with statistical normalization techniques, can be used to provide precise estimates of the number of people congregating in a cluster of city blocks during

market hours, allowing new developing markets to be detected.

CDR can also facilitate analysis of existing markets—for instance to detect overcrowding. While shopping at an open-air market can be part of the cultural lifeblood of a city, it comes with negative external cities like traffic/pedestrian congestion, and the geographic extent of existing markets may be misaligned with the actual number of shoppers frequenting it, leading to long lines and below potential sales volume. By counting the number of people in a market district during business hours, it is possible to derive geographic density measures (how many people per vendor) and identify markets with excess or insufficient sprawl. A market with routinely high density of CDR activity could be expanded to new streets—allowing for more vendors with more space to serve consumers—or a new market could be established nearby to draw shoppers to a less crowded space. Time-dependent analytics for each market could also be used to inform vendors as to shopper routines, allowing them to optimize opening/closing hours and staffing levels throughout the day.

Social Impact. Open-air markets are central to life in many cities. However, they can cause congestion, pollution, and litter in areas where they are not monitored and regulated. Additionally, open-air markets in the developing world are a major vector of transmission for communicable diseases like COVID-19, making overcrowding in these locations a major public health concern. CDR based analytics can help identify over- or undercrowded marketplaces, informing municipal planners and public health officials of opportunities for geographic expansion or contraction. Such data can also improve business for vendors by reducing crowding that inhibits sales volume, and

https://www.icirnigeria.org/report-how-open-air-markets-are-a-weak-link-in-the-fight-against-covid-19/

allowing them to adjust business practices to expected crowd sizes at different points in time.

7.4 Tracking Work Related Movements During A Lockdown

COVID-19 changed the nature of daily life for billions of people across the world. Between 2020 and 2022, there were more than 600 million cases of COVID-19, sparing no nation and taking more than 6.5 million human lives. Governments throughout the world implemented a variety of policy measures in response to the pandemic, including requests or mandates to shelter at home, cross-state travel restrictions, requirements for workers to work remotely where possible, and individual-level contact tracing in order to track COVID infections and prevent community spread.

"Aggregate and anonymized CDR data can be used to track the effects of public health measures like lockdowns."

Many COVID responses, however, were poorly targeted due to a lack of data on the actual locations of dense population centers, and compliance at the individual or aggregate level was difficult to track in the absence of data about peoples' movements during the pandemic. Contact tracing required training and mobilizing new workforces to manually reach the people with whom an infected person had interacted [46], whereas automated data analysis systems could—with user consent—have more efficiently harvested mobile contact and geolocation data to automatically identify potential new infections.

Aggregate and anonymized CDR data can be used to track the effects of public health measures like lockdowns. In Nigeria, in response to surging COVID cases, the government announced a lockdown in May 2020 that included work-from home orders, restrictions on interstate travel, as well as a nationwide curfew₍₃₅₎.

While the lockdown began to be gradually lifted within two weeks, compliance with restrictions was never systematically measured during the lockdown itself [48], leading to widespread violation of lockdown orders, instances of police violence, and potentially premature lifting of restrictions [35]. A government lockdown policy could have been dynamically informed by CDR analyses of worker movements in different localities, allowing health agencies to target awareness campaigns and enforcement activities to areas were transiting between work and home locations rather than abiding by the lockdown.

Since the benefit of a lockdown in terms of reducing disease spread and freeing up hospital capacity depends heavily on how much of the population follows the rules, measuring the rate of noncompliance could also have been used to provide real-time information about the policy's effectiveness, and adjust lockdown duration. Individual-level CDR data, if used carefully, can also be beneficial during epidemics.

The Chinese government implemented a COVID-19 "close contact app" that would inform users if they had come into contact with a COVID case [43]. Solutions like this can empower governments and individuals to protect themselves and society in response to rapidly-spreading and evolving pandemics, and such systems may indeed be part of routine disease management in

https://ourworldindata.org/coronavirus

the future. But while efficient, digital contact tracing at scale requires an underlying architecture of government data collection/surveillance. Systems like these should always be pursued with a clear-eyed understanding of privacy violations that can emerge from bulk collection of individual data, and should only be implemented in the presence of robust democratic debate across all segments of society, with built-in mechanisms for users to contribute their data only with affirmative consent.

Social Impact. The fields of public health and epidemiology depend on data that can serve as measurement tools for assessing health and predicting health at the individual and aggregate level, in service of building systems that can improve health for populations around the world. CDR data can help public health agencies access population-level indicators of mobile contact frequency and mobility that can be especially important during moments of crisis and disease spread. Paired with affirmative consent, CDR data can also be used as part of an automated contact tracing toolkit that empowers mobile users to protect their health during epidemics.



8 Use Cases: Health and Society

8.1 Improved Disaster Response
Management using Mobile Phone Network
Data — Human Mobility Pattern Tracking
in Northern Nigeria Using CDR

In recent years, Nigeria has experienced frenzy of organized violence to a degree and intensity that has never been experienced before. The Northern part of the country has recorded countless attacks from bandits and terrorist groups.

In addition to the pain and trauma directly caused to the victims of sectarian violence, attacks against communities in Nigeria has led to a displacement of those fleeing crisis regions, as individuals and families move to other parts of the country or cross national borders seeking to escape ethnic and religious persecution.

According to the United Nations Refugee Agency, violence in northern Nigeria has resulted in some 300,000 Nigerian refugees, with an additional 2.1 million people internally displaced within Nigeria. There are unique challenges and opportunities associated with CDR-informed humanitarian response, and a number of implemented systems have already been used in this context [27]. Flowminder18, a non-profit that monitors crisis related migration trends using a combination of cell data and surveys, deployed CDR to track 138,560

mobile SIM cards in Haiti after the 2010 earthquake, identifying regions at increased risk of subsequent cholera outbreak due to migration, with results that were more accurate than government estimates, especially during the period of diminished state capacity following a humanitarian disaster (13). Flowminder has likewise been deployed during the aftermath of the 2015 Nepal earthquake19, to study climate change-related migration in Bangladesh 20, and has been integrated into Mozambique's disaster response and management system to support real-time operational planning, resource allocation, and decision-making for the National Institute of Disaster Management (INGC)². A 2012 study also used CDR data to study displacement during the Rwandan genocide in 2005-2009, mapping trajectories for 1.5 million Rwandans over four years, overcoming the inherent difficulties of surveying highly mobile populations in crisis regions through data science not CDR can be used to provide a deep understanding of population mobility during periods of violence and insurgency. Insights from CDR can be used to inform evacuation plans by revealing where people in need of evacuation are located.

Beyond preparatory modeling, real-time data sharing agreements between mobile providers and governments/NGOs can be used for proactive crisis management by identifying regions where people are fleeing violence and the regions to which they are relocating, allowing for disaster response agencies to

allocate resources accordingly. Dynamic trajectory mapping can also be used to identify major thorough fares for migration, allowing security agencies to mobilize resources to protect populations that are in transit.

"CDR can be used to provide a deep understanding of population mobility during periods of violence and insurgency."

Social Impact. Millions of Nigerians have been displaced as a result of insurgency-related violence in recent years, leading to a colossal humanitarian crisis. The government's ongoing relief efforts could benefit tremendously from incorporating CDR-based migration modeling into existing disaster response into the existing disaster response architecture, ensuring that refugees are met with food, shelter, and social services that help them overcome the trauma of having survived violence. Government agencies such as the Ministry of Humanitarian Affairs and the National Emergency Management Agency as well as Humanitarian and Relief NGOs and foreign aid organizations could build CDR analysis in as a permanent feature of humanitarian response systems, allowing future emergency responses to be swift and well-targeted.

8.2 Mental Health State using Insights from Call Detail Records

Human beings are social creatures. The amount of social interaction a person engages in is both a reflection and a determinant of mental health. People experiencing symptoms of depression or anxiety often withdraw from

their social networks—a key diagnostic criterion for these conditions [33]—leading to a vicious cycle of isolation and loss of a community of support, heightening mental health crises. People who engage in regular social interaction have been shown to have better mental health [23], and promoting social engagement by strong individual and collective senses of belonging, social identity, and personal fulfillment [44] is a major public health goal.

CDR can be used to gather data on patterns of social engagement at both the individual and the aggregate level. The level of calling and texting activity reflected in a mobile user's call records can be used to profile the amount and diversity of their socialization patterns.

Although mobile data only captures part of a person's socialization patterns—which include face-to-face interactions both at work and at home—people increasingly communicate over the phone and online, and research has shown that phone activity measures not just of volume of phone activity, but also the nature of phone activity (e.g. how often are they callers versus recipient of calls?) are predictors of depression in some populations, with interactive users who both make and receive calls having the lowest risk [o]. Further metrics can be designed to measure predictors of depression on the basis of a user's local social network graph. For instance, research indicates that people with depression tend to socialize with a small number of people as compared with the mentally healthy, who tend to engage in a greater diversity of individual and group interactions [23]. Measures like communication network entropy—the frequency with which a person exchanges calls or texts with just one node in their social graph versus distributing their communication activity across a large number of others—can serve as indicators of the diversity of social engagement, providing unexplored features likely to be predictive of mental health in CDR.

"Research has shown that phone activity measures not just of volume of phone activity, but also the nature of phone activity are predictors of depression in some populations."

Beyond providing mental illness detection services for individuals, CDR can also inform studies and interventions at the aggregate level by measuring social engagement in geographically localized communities using anonymized call records. Rural isolation is well-known to correlate with mental illness [45], though geographical distance can be overcome by social engagement via technology. CDR can be used to detect communities that are isolated not just geographically but that have high rates of social isolation, providing targets for public health intervention.

Social Impact. CDR-derived metrics predict depression at the individual level, and that the community level can be used by groups like the Mental Health Foundation Nigeria, Nigerian Medical Association, and World Health Organization to discover community-level social isolation and design programs to promote mental health by building robust social support systems.

8.3 Estimating the Extent of Air Pollution using Call Detail Records

We do not all breathe the same air. Pollution levels can vary dramatically even across very small geographic distances. A city block adjacent to a busy highway is exposed to much more air toxins than a nearby neighborhood adjacent to a public park. Air pollution is of particular concern in Nigeria, which has the world's highest number of pollution-related child pneumonia deaths in the world [1]. In Lagos State, home to 24 million people, has particulate pollution levels that are 6.8 times the WHO-recommended level, causing 11,200 premature deaths and \$2.1 billion in 2018 according to the World Bank [28]. Air pollution intensifies with urban growth, and Lagos is projected to become the world's largest city by 2100, making it all the more important for public officials to identify pollution zones and act to reduce levels of air and water toxins emitted by industrial and transit activity.

Mobile phone data have been used in the past to refine air pollution exposure estimates by incorporating individual-level mobility data into estimates of how much pollution a person encounters in the course of their daily activities. Estimates of individual pollution risk are usually based on a person's home location and measured pollution levels at that address. However, people can experience differing levels of pollution in the workplace (e.g. if they commute from the suburbs into a factory) and as they are in transit (e.g. if stuck in a highway traffic jam on their way to work). Yu et. al. (2020) [55] used detailed mobility information for more than 310,000 people in Shenzhen (pop. 17 million) to estimate individual pollution exposure using 35 million CDRbased location samples. They found that standard methods of assessing individual pollution risk underestimated transit-related particulate exposure especially during transit periods such as rush hour, while overestimating home-location ozone exposure.

Methods for interpolating individual trajectories rather than discretely sampling stay-points (see Section 4.1) could allow for even more accurate estimates, which could be used to build tools allowing mobile users to get detailed information about the pollution they encounter in the course of their day, proposing alternate routes they could use to avoid pollution-heavy areas. Public health agencies could use CDR for detailed analysis of commuter town pollution exposure, directing health resources and public awareness campaigns to communities whose pollution levels would be underestimated if only assessed based on their home location without including transit data.

Social Impact. Pollution is a major cause of ill health and death, particularly in Nigeria. Public health efforts to curb pollution exposure could include individual-level modeling tools that give mobile users detailed information about the pollution they encounter throughout their day, empowering them to make choices about transit routes that would reduce their exposure. Population-level CDR analyses can be used to by NGOs and ministries of health identify at risk communities, and also provide dynamic population estimates useful for estimating the amount of pollution produced by human activity in a given area.



- [1] Nigeria has highest number of air pollution-related child pneumonia deaths in the world, Nov. 2021.
- [2] World toilet day sees limited progress in Nigerians' access to sanitation services. Nov. 2021.
- [3] Gps accuracy, Mar 2022.
- [4] Ameh Ochojila Abuja. Stakeholders advocate enactment of laws to combat open defecation in Nigeria, Sept. 2022.
- [5] Mark Quaye Affum. The effect of internet on students studies: A review. Library Philosophy and Practice, February 2022.
- [6] Emmanuel Akinwotu. In Lagos, the traffic jams can add four hours to your commute, Sep 2015.
- [7] O.R. Akpomrere and O. Nyorere. Bus Rapid Transit (BRT) and railway transport system in Lagos: Geographic information system approach. International Journal of Economic Development Research and Investment, 4, April 2013.
- [8] Nour Raeef Al-Molhem, Yasser Rahal, and Mustapha Dakkak. Social network analysis in telecom data. Journal of Big Data, 6, November 2019.
- [9] Timothée Aubourg, Jacques Demongeot, Félix Renard, Hervé Provost, and Nicolas Vuillerme. Association between social asymmetry and depression in older adults: A phone call detail records analysis. Nature Scientific Reports, 9, 2019.
- [10] Oluleke Bamodu. Application of mobile technology in waste collection. International Journal of Applied Sciences (IJAS), 4, 2013.
- [11] Edward Barbour, Carlos Cerezo Davila, Siddharth Gupta, Christoph Reinhart, Jasleen Kaur, and Marta C González. Planning for sustainable cities by estimating building occupancy with mobile phones. Nature Communications, 10, August 2019.

- [12] Daniel Gashaney Belay, Melaku Hunie Asratie, Fantu Mamo Aragaw, Nuhamin Tesfa Tsega, Mastewal Endalew, and Mooges Gashaw. Opendefecation practice and its determinants among households in sub-saharan africa: pooled prevalence and multilevel analysis of 33 sub-saharan Africa countries demographic and health survey. Tropical Medicine and Health, 50, March 2022.
- [13] Linus Bengtsson, Xin Lu, Anna Thorson, Richard Garfield, and Johanvon Schreeb. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earth quakegeospatial study in Haiti. PLOS Medicine, 8, August 2011.
- [14] C.B. Bhattacharya. When customers are members: Customer retention in paid membership contexts. Journal of the Academy of Marketing Science, 25, 1998.
- [15] Nicola Bicocchi and Marco Mamei. Investigating ride sharing opportunities through mobility data analysis. Pervasive and Mobile Computing, 14:83—94, 2014. Special Issue on Pervasive Education Special Issue on The Social Car: Socially-inspired Mechanisms for Future Mobility Services.
- [16] Joshua E. Blumenstock. Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda. Information Technology for Development, 18:107—125, Feb. 2012.
- [17] Paige M. Boshell. The power of place: Geolocation tracking and privacy, Mar 2019.
- [18] Federico Botta. Quantifying the differences in call detail records. Royal Society Open Science, 8, June 2021.
- [19] Ani Casimir, Emma Omeh, and Chinedu Ike. Electoral fraud in Nigeria: A philosophical evaluation of the framework of electoral violence. Open Journal of Political Science, 3, 2013.
- [20] Larry E. Daniel and Lars E. Daniel. Chapter 33 cellular system evidence and call detail records. In Larry E. Daniel and Lars E. Daniel, editors, Digital Forensics for Legal Professionals, pages 225—237. Syngress, Boston, 2012.

[21] Umar Lawal Dano, Abdul-Lateef Balogun, Ismaila Rimi Abubakar, and Yusuf Adedoyin Aina. Transformative urban governance: confronting urbanization challenges with geospatial technologies in Lagos, Nigeria. GeoJournal, 2019.

[22] Koustuv Dasgupta, Rahul Singh, Balaji Viswanathan, Dipanjan Chakraborty, Sougata Mukherjea, and Amit A. Nanavati. Social ties and their relevance to churn in telecom networks. In EDBT. March 2008.

[23] Timon Elmer and Christoph Stadtfeld. Depressive symptoms are associated with social isolation in face-to-face interaction networks. Nature Scientific Reports, 10, 2020.

[24] Dario Forte and Andrea de Donno. Chapter 10 - mobile network investigations. In Eoghan Casey, Cory Altheide, Christopher Daywalt, Andrea de Donno, Dario Forte, James O. Holley, Andy Johnston, Ronald van der Knijff, Anthony Kokocinski, Paul H. Luehr, Terrance Maguire, Ryan D. Pittman, Curtis W. Rose, Joseph J. Schwerha, Dave Shaver, and Jessica Reust Smith, editors, Handbook of Digital Forensics and Investigation, pages 517—557. Academic Press, San Diego, 2010.

[25] Ronnie Garrett. Analyzing the hidden patterns in call records. Law Enforcement Technology, 29:96—99, May 2002.

[26] Modupe Gbadeyanka. Open-air markets account for 97% of Nigeria's retail sales, July 2022.

[27] Talia Gertstle, Maya Hardimon, Galen Hunt, Iona Main, Liza Paudel, Daniel Pontón, Sri Harshita Rallabhandi, Kaitlin Roh, Maria Romano, Emily Tenenbom, and Julia Tuttle. Assessing the use of call detail records (CDR) for monitoring mobility and displacement. Technical report, United Nations on Migration/Princeton School of Public and International Affairs, Feb. 2021.

[28] World Bank Group. The cost of air pollution in lagos, Sept. 2020.

[29] Thoralf Gutierrez, Gautier Krings, and Vincent D. Blondel. Evaluating socio-economic state of a country analyzing airtime credit and mobile phone datasets, 2013.

[30] Corrie Hannah, Julia Davies, Rachel Green, Andrew Zimmer, Patrese Anderson, Jane Battersby, Kathy Baylis, Nupur Joshi, and Tom P. Evans. Persistence of open-air markets in the food systems of Africa's secondary cities. Cities, 124:103608, 2022.

[31] Michael V. Heinz, Natalie X. Thomas, Nhi D. Nguyen, Tess Z. Griffin, and Nicholas C. Jacobson. Technological advances in clinical assessment. In Comprehensive Clinical Psychology, volume 4, pages 301—320. Elsevier, 2 edition, 2022.

[32] Margo Hilbrecht, Bryan Smale, and Steven E Mock. Highway to health? commute time and well-being among Canadian adults. World Leisure Journal, 56:151—163, 2014.

[33] Sahar Hoteit, Stefano Secci, Stanislav Sobolevsky, Carlo Ratti, and Guy Pujolle. Estimating human trajectories and hotspots through mobile phone data. Computer Networks, 64.

[34] Georges V. Houngbonon, Erwan Le Quentrec, and Stefania Rubrichi. Accessto electricity and digital inclusion: evidence from mobile call details records. Nature Humanities & Social Sciences Communications, 8, July 2021.

[35] Ridwan Lanfre Ibrahim, Kazeem Bello Ajide, and Omokanmi Olatunde Julius. Easing of lockdown measures in Nigeria: Implications for the healthcare system. Health Policy Technology, 9:399—404, Dec. 2020.

[36] Lisa Ivers, Stefano Niavas, Chris Mitchell, Zineb Sqalli, and Omar Frikha. The future of traditional retail in Africa, June 2022.

[37] Gabi Kedma, Mordehai Guri, Tom Sela, and Yuval Elovici. Analyzing users' web surfing patterns to trace terrorists and criminals. In IEEE International Conference on Intelligence and Security Informatics, June 2013.

- [38] Saiqa Khan, Firozz Ansari, Hamza Azmi Dhalvelkar, and Sabiqua Computer. Criminal investigation using call data records (cdr) through big data technology. In International Conference on Nascent Technologies in the Engineering Field (ICNTE), 2017.
- [39] Elizabeth Kneebone and Natalie Holmes. The growing distance between people and jobs in metropolitan America, Mar 2015.
- [40] Rainer Kujala, Talayeh Aledavood, and Jari Saramäki. Estimation and monitoring of city-to-city travel times using call records. EJP Data Science, 5, March 2016.
- [41] Rosa Lavelle-Hill, John Harvey, Gavin Smith, Anjali Mazumder, Madeleine Ellis, Kelefa Mwantimwa, and James Goulding. Using mobile money data and call detail records to explore the risks of urban migration in Tanzania. EJP Data Science, 11, May 22.
- [42] Joe Lee. Internet usage effects on educational attainment: Evidence of benefits. Technical report, Milken Institute Center for Jobs and Human Capital, July 2017.
- [43] Donna Lu. China uses mass surveillance tech to fight spread of coronavirus. New Scientist, 245:7, Feb. 2020.
- [44] Mengyun Luo, Ding Ding, Adrian Bauman, Joel Negin, and Philayrath Phongsavan. Social engagement pattern, health behaviors and subjective well-being of older adults: an international perspective using who-sage survey data. BMC Public Health, 20, 2020.
- [45] Dawn A. Morales, Crystal L. Barksdale, and Andrea C. Beckel-Mitchener. A call to action to address rural mental health disparities. Journal of Clinical and Translational Science, 4:463—467, 2020.

Al Startup Hub & Research Lab 33, Queens Street, Alagomeji, Yaba, Lagos.

+234 906 200 0354 www.datasciencenigeria.org