BIG DATA *and Thriving CITIES*

Innovations in analytics to build sustainable, resilient, equitable and livable urban spaces

WORLD BANK GROUP

CONTENTS

INTRODUCTION

Economic growth and urbanization are complementary and critical to poverty reduction in low- and middle-income countries. With 450 million new urban dwellers expected between 2010 and 2040, the typical African city will double its population and many new cities will be built[1]. For many developing countries this presents an opportunity to leapfrog development cycles, catch up and even surpass some of the 21st-Century hubs of high-income nations.

Rapid urbanization is frequently unplanned, as cities struggle to keep up with the infrastructure needs of fast-growing populations. Most rapidly expanding cities also lack the data needed even to monitor growth, let alone to guide it. Newly available technologies, techniques and intellectual capital provide a huge opportunity to help close this gap. These new approaches are already being deployed worldwide in order to improve insights into urbanization and for data-driven decision making. The recent global diffusion of new technologies, combined with the use of big data analytics, can help policymakers promote the effective development of future cities that provide living and work environments in which citizens can thrive[2].

In particular, innovative applications of geo-spatial and sensing technologies and the penetration of mobile phone technology are providing unprecedented data collection

1. https://openknowledge.worldbank.org/bitstream/handle/10986/16657/815460WP0Afric00Box379851B00PUBLIC0.txt?sequence=2
2. The living environment determines the quality of housing, access to public goods, opportunities for recreation and socialization, and life prospects of children. The work environment determines the opportunities that exist for education, jobs and income generation.

opportunities. This data can be analyzed for many purposes, including tracking population and mobility, private sector investment, and transparency in federal and local government. It can be visualized accurately and in real-time, helping policymakers to make informed investments in infrastructure to meet both short- and long-term needs, while minimizing opportunity costs.

To help development practitioners within and beyond the World Bank take advantage of these trends, this brief profiles a sample of big data applications to support improved urban development in low- and middle-income countries. It also cites potential opportunities for big data analytics to help developing nations achieve sustainable urban growth, while reducing the economic differential with high-income countries.

WHAT IS BIG DATA IN A DEVELOPMENT CONTEXT?

Big data is an umbrella term used to describe the constantly increasing flows of data emitted from connected individuals and things, as well as a new generation of approaches being used to deliver insight and value from these data flows. It is said that more data has been generated in the past two years alone than in all previous years combined. Big Data can be defined as high-volume, high-velocity and high-variety datasets that can be analyzed to identify and understand previously unknown patterns, trends and associations.

While most of the attention given to big data has focused on high-income countries, the rapid diffusion of technologies such as the internet, cellphones, ground sensors, drones and satellites – to name a few – is also driving big data innovation in low- and middle-income countries. And while data flows in the developing world are typically smaller and less diverse than in the developed world, they still present incredible opportunities for data scientists, economists and statisticians to use big data to enhance or supplement traditional analytical approaches.

Developing world big datasets still maintain many of the unique characteristics that make big data different from traditional datasets. They include the comparatively large volume of data, its varied and unconventional sources, and the relative speed with which it accumulates. Such characteristics enable analysts to access radically new insights and understand phenomena that traditional data collection systems cannot offer. Unlike previous sources of development data, such as household surveys, which address specific research questions, big data is usually produced in the course of some other activity (such as making a cellphone call). This, along with the size and complexity of some datasets, requires different research methods. Big data analytics is the emerging set of tools and methods to manage and analyze this explosive growth of digital information. It includes data science methods such as machine learning, predictive analytics and visualization. These methods offer significant opportunities to draw on real-time information to address development challenges.

The spread of analytics expertise, open-source software and low-cost analytics packages means that big data will become increasingly indispensable in helping low- and middle-income countries analyze trends and develop policy. The potential of big data in developing nations will only grow as they continue to digitize fast.

BIG DATA IN URBAN DEVELOPMENT

Big data and urban spaces are closely connected. The concentration of people and data-producing assets such as phones, traffic and security cameras means cities naturally generate massive amounts of data. Formal government data on factors such as health, the economy or the environment are also prolific in cities. In lower-income countries, capacities for data collection and analysis are often most available in large cities, where technology and state services are most developed. This abundance of data means big data analytics has strong potential to help generate new approaches and solutions to challenges in urban development.

There are, however, particular challenges to using big data in low- and middle-income cities. Analytics alone cannot create better cities. It can inform the development and implementation of solutions, but government policies, investment, financing, local support and capable institutions are also essential. Less developed urban environments may lack the technological resources to collect data, such as closed-circuit TV, smartphones or traffic sensors. Biases can be introduced when data is not representative – for example, the lower end of income distribution can be missed from data collected through cellphones. While not an important issue in higher-income countries, bias is likely to be a serious challenge in lower-income contexts. In addition, the extensive informal exchange networks that are vital to urban economies in low- and middle-income countries are extremely difficult to track, even with the latest technology. This suggests that big data approaches in such contexts cannot rely on a hardwired urban data collection infrastructure, such as traffic monitors, or high-tech tools such as GPS-enabled smartphones. Rather, they may be best applied to phenomena that can be tracked using cheaper passive systems, such as low-cost cameras, satellite imaging, cellphones and crowdsourcing. National restrictions on privacy or the use of data may also affect the extent to which big data can be used in specific countries.

Despite these potential hurdles, numerous innovative big data programs have already been created or implemented in low- and middle-income countries. Many are on a par with the developed world and may even present models for high-income cities to follow.

PROFILING BIG DATA APPROACHES

The increased availability and reduced cost of technology is allowing low- and middle-income countries to leapfrog many of the steps taken in high-income nations to develop a robust technological infrastructure. As digital progress continues, many more big data approaches will be harnessed for urban development in lower-income countries, at a faster rate than before.

Numerous examples show that such approaches offer valuable new ways of understanding emergent events and trends that are difficult to measure. This brief offers four key case studies, and many more examples, as a taster of the potential of big data analytics in low- and middle-income urban contexts. These cover different implementation timeframes, from already being deployed to expected readiness within the next three years. They focus on how the outputs of big data analytics can be used to improve policy in five focus areas of the World Bank's urban work: Green Cities, Inclusive Cities, Systems and Governance, Resilient Cities and Competitive Cities.

The brief's conclusion looks at big data's role in shaping the cities of the future, highlighting exciting technologies and techniques that could emerge in low- and middle-income countries over the coming decade. All these big data approaches are applicable in a wide variety of countries and contexts. They could achieve improvements as diverse as informing better urban planning, helping beat congestion and crime, or promoting financial services and efficient tax collection. Together they showcase new possibilities for innovation in the quest for functioning, resilient and sustainable cities that enhance their residents' lives.

KEY TO CASE-STUDY TIMEFRAMES

KEY TO CASE-STUDY POLICY AREAS

World Bank staff can find tools, training and knowledge on big data on the intranet at //bigdata

CASE STUDY

Using Geospatial Data to Track Changes in Urbanization

Resilient Cities

Inclusive Cities

NEAR-TERM

HIGHLIGHTS

- Analysis of Earth observation (EO) big data from satellites and sensors can help stakeholders track and understand urban development over time.
- Working with providers of geospatial data, the World Bank has carried out analytics to measure the qualitative and quantitative aspects of urban transformation, such as the distribution and density of urban sprawl, changes in land use and the growth rate of built-up areas. This allows analysts to begin to understand the drivers of land consumption.
- Combining EO data with other data on population and growth reveals dramatic insights on the overall economic viability, inclusivity, resilience, sustainability and quality of life in urban areas. This enables stakeholders and policymakers to develop and implement informed policies in response, to create thriving cities of the future.

PUBLIC POLICY USES FOR EO BIG DATA

- Track the growth of urban areas and understand economic drivers.
- Evaluate the current state of amenities and identify opportunities and priorities for developing dynamic, equitable, sustainable and resilient cities.
- Underpin smart policymaking to promote optimal spatial and transportation links between jobs, affordable property, health and education services, and recreational areas.

Practice Areas:
Resilient Cities, Inclusive Cities

Countries Involved:
Various in South Asia, East Asia and Africa

Data Types:
Remote sensing, satellite imagery

The World Bank has been using big data to track and study changes in urbanization in low- and middle-income countries, to help ensure that it is sustainable, equitable and supports economic growth. As the world's population becomes more urbanized, reducing unemployment and promoting sustainability and resilience in urban economies is vital. However, much urbanization is not well planned or managed in developing countries, resulting in cities which have grown rapidly but lack critical infrastructure and are unable to take advantage of economies of scale. For citizens, this means poor transport and services, weak links between job opportunities and the workforce, increased inequality, low resilience to shocks and a lack of sustainability.

Successful urbanization depends on the coordination of three distinct but interdependent processes: public investment in infrastructure, private investment in productive capital and household investment in housing[3]. However, the speed of development and the lack of information available to each set of actors often prevents this coordination. Big data can play an important role filling this gap and facilitating improved coordination. It is especially valuable where the speed and pace of urbanization outstrip the authorities' ability to understand how cities are growing and changing. Without this baseline information, policymakers cannot meet the challenges and opportunities of urbanization through properly informed decisions.

Visualizing urban landscapes

In South and East Asia, the World Bank has explored the patterns, consequences and policy implications of cities' spatial development by drawing on the increasing availability of spatial data and developments in analytics. Satellite or Earth observation (EO) data can deliver quality results to measure urban growth over a wide range of spatial and temporal scales, particularly when combined with data from other sources. The resulting digital urban maps provide an up-to-date, accurate and cost-effective resource to help national, regional and city governments understand the nature of urban development and make informed decisions. EO datasets allow harmonized and standardized measurements, enabling spatially and temporally consistent comparisons, as well as global assessment. Such data is particularly important for monitoring and understanding the evolution of cities – for example, allowing officials to see when built-up areas spill across formal administrative boundaries. This indicates the need to work with adjoining administrative areas on issues such as connective infrastructure (roads, water mains) or collecting garbage.

Drawing on analysis of EO data, the World Bank has built a database that describes

3. Source: Building African Cities that Work: A study on the Spatial Development of African Cities (P148736). World Bank concept note, November 13, 2013

the speed, magnitude and spatial form of urbanization. Using these data, Bank teams have examined the drivers and impacts of the nature of urbanization and how the urban landscape has evolved to its current state. This provides a baseline from which to understand the effects of policy change and identify priorities for new initiatives. In particular, the Bank focused on exploring the institutional frameworks for urban management (such as mechanisms to coordinate service delivery across administrative jurisdictions), investment (in transport, services and other network infrastructure) and regulation (such as land use, zoning and pricing of services).

In 2008, in close collaboration with the European Space Agency (ESA), the World Bank launched the "Earth Observation for Development" initiative. This provides data on urban and other trends in areas where data are traditionally scarce and often unreliable. Such information can be used to establish project baselines against which progress can be gauged, mitigation measures determined and high-priority issues identified. The initiative focuses on a number of areas, including urban development and related fields such as disaster risk management, the environment, water and energy. To facilitate greater collaboration towards these objectives between policymakers and other development stakeholders, the Bank also developed the Platform for Urban Management and Analysis (PUMA)[4]. This tool allows users with no prior GIS experience to access, analyze and share urban spatial data in an interactive and customizable way.

These initiatives have led to more than 30 World Bank technical assistance projects delivered to national or city partners or due for completion in the 2008-18 period. The results have been used to maintain and strengthen policy dialogues, as well as to guide the design of new projects. They have led to highly specialized big data mapping products and monitoring systems that leverage EO data for South Asian cities – for example, mapping urban extent and analysis of the internal spatial structures of different cities, or monitoring land subsidence in a city with a high reliance on tube wells for water, lowering the water table. The World Bank-ESA partnership has been expanded until 2018, with Urban Development among its top three priority themes. It aims to provide a systematic source of development information, so that stakeholders can draw on state-of-the art EO capabilities to develop best practices and sustainability plans.

Mapping the megacities

Under the World Bank's South Asia Megacities Improvement Program, EO big data was used to analyze 20 years of urban expansion in the metropolitan areas of Delhi, Mumbai and Dhaka. These data enabled measurement of the qualitative and quantitative aspects of urban transformation, such as the distribution and density of urban sprawl, changes in urban land use and the growth rate of built-up areas. Using this information, analysts can see how informal settlements grow outside the cities' administrative boundaries, and can begin to understand the drivers of land consumption.

4. Available at http://puma.worldbank.org/

The analysis revealed important insights into land cover and use in the three cities (see Figure 1), revealing the percentage of land taken by residential build-up, industrial build-up, agriculture, natural or semi-natural vegetation and forest. This helps city planners and development stakeholders understand existing requirements and plan for future needs.
In Delhi, for example, the maps show the urban sprawl is accelerated by industrial development. This mainly took place between 2003-10, although a significant increase in construction sites indicates that it will continue in the future – and must therefore be planned for.

Digitized spatial data allows analysis at different administrative levels: metropolitan, city, district or sub-district, as well as other non-administrative units. Such datasets allow flexible aggregation, for instance, showing the proportion of sprawl by district, its distribution and density, class evolution within urban areas and the drivers of urban change. (If housing comes before roads, for example, this indicates informal and incremental city building. If roads come first, development is formal.) Combined with environmental or socio-economic data, the data can provide information concerning the ratio of population growth to urban growth, and can measure indicators such as compactness (as a function of city density), the ratio of green space to citizens, and the proximity and accessibility of green areas.

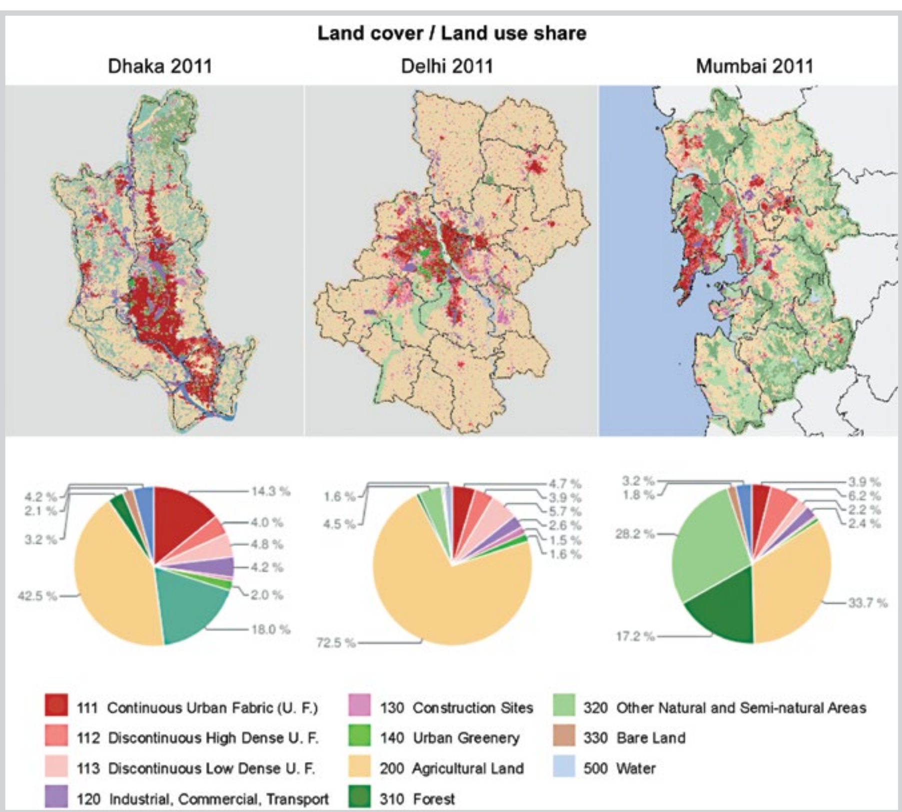


Figure 1: Sample visualizations from the South Asia geospatial analysis

In 2014, the World Bank used satellite imagery and demographic data to measure expansion and population change between 2000-10 in East Asian urban areas of 100,000 people or more. Analysts used change-detection methods that draw on satellite data from the Landsat remote sensing project operated by the US Geological Survey and the National Aeronautics and Space Administration (NASA). These maps rely on a geophysical definition of built-up areas as landscape units with more than 50 percent coverage of non-vegetative, human-constructed elements. These areas were combined with the AsiaPop map, from the world mapping project[5]. The refined land cover datasets were then combined with population density information derived from census data, and used to disaggregate population counts to a grid of 100 meters squared. This approach has allowed the Bank to understand the entire region so as to establish systematically where urbanization is occurring, to what degree and how quickly. This highlights the responses required from stakeholders, such as meeting needs for services (water, sanitation or transport) and regulation. Analysts also quantified the relationship between urbanization, income growth and inequality.

The approach provided critical information on the growth of urbanization in East Asia. For example, by 2010, the region had 869 urban areas with more than 100,000 people (600 of these in China). Populations in these cities had grown dramatically, lifting the region's new urban population to approximately 200 million people. However, despite this growth, only 36 percent of the region's total population lives in urban areas. This suggests that the trend in urban growth is likely to continue for many decades. Lower-middle-income countries such as Indonesia, the Philippines and Vietnam had the fastest urban population growth, whereas upper-middle-income countries such as China had the fastest spatial growth (hence most urban areas outside China became denser, while the density of many Chinese urban areas declined). Most of this growth occurred in small and medium-sized cities rather than the megacities. Unsurprisingly, urban density was high, and increased through the period.

Findings from EO big data can help coordination between public investment in infrastructure, private investment in productive capital and household investment in housing. They can allow policymakers to promote optimal spatial and transportation links between jobs, affordable housing and business units, health and education services and recreational areas. These insights can also be used to help support recent rural-to-urban migrants, bringing them to the attention of city authorities and ensuring that rapid urbanization is inclusive. As EO-based big data techniques spread across Africa and other developing regions, and are refined and adapted, they will provide valuable tools and insights to policymakers, and even greater benefits for the citizens of the future.

Creating livable cities

Further EO big data approaches are also helping drive sustainable urban development. Research into using high-resolution satellite data for poverty mapping in Sri Lanka draws on emerging techniques that can profile fast-

5. http://www.worldpop.org.uk

changing urban areas in near-real-time. These techniques can identify built-up area, building and car density, and types of roofing and road. Using open-source image-processing algorithms, they can even calculate whether buildings are more rectangular or have more chaotic angles (indicating higher poverty) and construct poverty indicators such as the percentage of paved roads in an area. This helps stakeholders target their interventions precisely where they are most needed.

In Kosovo, after a period of chaotic urban expansion, geo-spatial data collected by drones has been used to secure property rights quickly and cheaply – vital to household security and economic growth. The drones record imagery which is processed into high-resolution orthophotographs (aerial photographs corrected to have the same lack of distortion as a map) in a fraction of the time of conventional aerial surveys. The orthophotos are used to gather property boundary information from local residents, which can be used for formal property registration. In the fast-growing city of Ferizaj, this approach was used to support a government program for unregistered land owners to legalize their property rights. The data from the drone flights were processed in 24 hours using two local high-end desktop computers, resulting in orthophoto maps from which land owners could easily identify their property. The initiative can be scaled up globally, especially to secure land rights in low-income countries.

All these projects demonstrate that analysis of EO big data can be a significant tool for managing urban development in low- and middle-income countries. It can measure, baseline and track the growth of urban areas, and highlight the drivers of economic growth. This allows policymakers and stakeholders to better understand the factors leading to inefficiencies and inequality in urban areas, and to develop informed policies in response. They can also build resilience into urban environments, so that residents, institutions, businesses and systems can adapt to chronic stresses or acute shocks. And they can create livable cities that fulfill their residents' needs.

An Earth observation project shows a successful approach to tracking changes in urbanization using geospatial data

HIGHLIGHTS

- New cloud-based computational platforms such as Google Earth Engine (GEE) enable urbanization to be monitored using multi-spectral imagery from multiple satellites over extended time periods.
- Current and historic satellite data can be ground-truthed by manually labeling areas as "built-up" or "not built-up". A machine-learning algorithm allows for the conversion of these data into highly-accurate classification of

land for a particular area – in this case, cities in India.

- This allows urbanization to be measured with a high degree of geographic precision and close to real time, which will transform public policy design to help tackle the economic and environmental challenges of urbanization.

POTENTIAL USES FOR MONITORING DATA FROM MULTIPLE SATELLITES OVER TIME

- Accurate, near-real-time detection of urban areas – even across large-scale regions with diverse land cover – allowing urbanization to be traced and understood
- Urban planning based on unprecedented levels of information about past, current and predicted urban growth
- Understanding of the ecological, environmental, social and economic impacts of urbanization, from land classification data combined with other geographic data

While urbanization in rapidly growing nations is helping lift hundreds of millions of people out of poverty, it is also creating immense societal challenges. It is expanding greenhouse gas emissions, destabilizing fragile ecosystems, and creating new demands on education, health, and transportation infrastructure. Despite the importance of understanding the drivers of urban growth, it is still not possible to quantify the magnitude and pace of urbanization at a global scale. Standard empirical approaches use data from household surveys, which are costly to carry out, produced infrequently and subject to often-severe measurement problems. Reliable and up-to-date data on urbanization – particularly from developing countries – remain scarce.

The coming revolution in geospatial data holds the potential to transform the way in which we study cities. As satellite imagery at ever-improving spatial and temporal resolutions becomes available, new approaches to machine learning are being developed that convert these images into meaningful information about the nature and pace of change in urban landscapes. Research on land use is rapidly shifting towards remote-sensing methods designed to capture urban features as they are observed in terrestrial Earth observation data.

Despite the promise of satellite imagery for urban analysis, current information on land use is still subject to many drawbacks. Existing satellite-based classifications of urban areas cover limited geographic extents and time periods, and frequently disagree in terms of the size and shape of particular cities. Progress is further inhibited by lack of large datasets that give the "ground truth" regarding urbanization, which are essential for validating the micro-detailed maps of urban areas that remote-sensing methods produce. These deficiencies mean that urbanization still cannot be tracked with a high degree of precision across space and time.

Global-scale analysis of satellite imagery

As powerful new cloud-based computational platforms become available to the research community, it is becoming feasible to monitor urbanization using multi-spectral imagery from multiple satellites over an extended time-period. One such platform is Google Earth Engine (GEE). GEE leverages cloud-computational services for planetary-scale analysis and consists of petabytes of geospatial and tabular data. This includes a full archive of scenes from the US remote sensing project Landsat (as well as other satellites), together with a JavaScript, Python-based application programming interface (API), and algorithms for supervised and unsupervised image classification.

Recent World Bank research demonstrates the applicability of GEE for studying urban areas at scale. It provides, for the first time, reliable and comprehensive open-source, ground-truth data for supervised image classification that delineates urban areas – in this case, in India. As a large, geographically diverse nation undergoing a rapid urban transition, India represents an ideal context in which to illustrate the applicability of new approaches for mapping urban expanse. The team leveraged the computational power of GEE and its full Landsat archive to introduce a practical and adaptable procedure for analysis of urban areas at a global scale.

The dataset consists of 21,030 polygons that were manually labeled as "built-up" urban areas or "not built-up" areas. To generate a machine-learning algorithm that allows for the conversion of these ground-truth data into a classification for the country as a whole, the team assessed alternative supervised classifiers (Random Forest, Classification and Regression Tree (CART) and Support Vector Machine) and examined the effects of various inputs and class combinations on the performance of the classifiers. They proposed a methodology – "spatial k-fold cross-validation" – to evaluate the extent to which the classifiers

Figure 2: Classification of built-up areas (visualized in red) compared to raw satellite images in three regions in India (Classifier: Random Forest with 10 trees; Input: Landsat 8).

could be generalized in spatial terms, and their performance in a large and geographically heterogeneous context.

Accurate mapping across space and time

The team found that their ground-truth dataset can be used in GEE to produce high-quality maps of built-up areas in India, across space and time. Their classification captures the fabric of built-up urban areas, as well as the fine boundaries between cities and their peripheries.

In validating the classification results, the team showed that when used with standard classifiers available in GEE, they achieve a high overall accuracy rate of around 87 percent in identifying built-up areas for grid cells with a dimension of 30 meters. Of the three types of classifiers examined, Random Forest achieved the best performance (a balanced accuracy rate of 80 percent). However, the performance of the classifiers strongly improves as the size of the training dataset increases (especially with CART). With this methodology, which is designed to evaluate the spatial generalizability of classifiers, the team showed that the classifiers also perform well when the training examples are sampled from areas with heterogeneous land cover (such as a mix of dense vegetation and bare ground).

As inputs for the classifiers, the team used Landsat 7 and Landsat 8, launched in 1999 and 2013 respectively. Although Landsat 7 is of a lower spectral resolution than Landsat 8, it allows for a longer time horizon over which to study urbanization. The project demonstrated that the addition of two pixel-level indices to Landsat 7 as inputs to the classifiers – the Normalized Difference Vegetation Index and the Normalized Difference Built-up Index – improves classifier performance so it approaches that when Landsat 8 is used for the input. This shows that the classifiers can be used to detect urban areas in historic data. When the extent of urban areas in 2000 was mapped in this way, the team found a high overall classification accuracy of 86 percent.

Towards precise, real-time urban measurement

Urbanization is a fundamental driver of economic growth in the 21st century. Although it implies massive productivity gains for an economy, it is also accompanied by congestion, pollution and heavy demands on public-sector resources. Understanding the ecological, environmental, social and economic impacts of these changes is essential for preserving a sustainable society.

As parallel computational platforms become accessible to researchers, it is now possible to expand urban research across space and time. In this study the team developed a large-scale geo-referenced dataset that was used to facilitate the detection of urban areas at a national level, and to provide a useful and reliable tool for temporal analysis of urban zones and their rural peripheries. The study highlights the potential of GEE for urban research and illustrates the applicability of the dataset for the detection of urban areas in a country with a large population and a diverse land cover. The methodology and the evaluation procedure are suitable for studies that analyze large-scale regions, and can easily be applied to other countries and contexts.

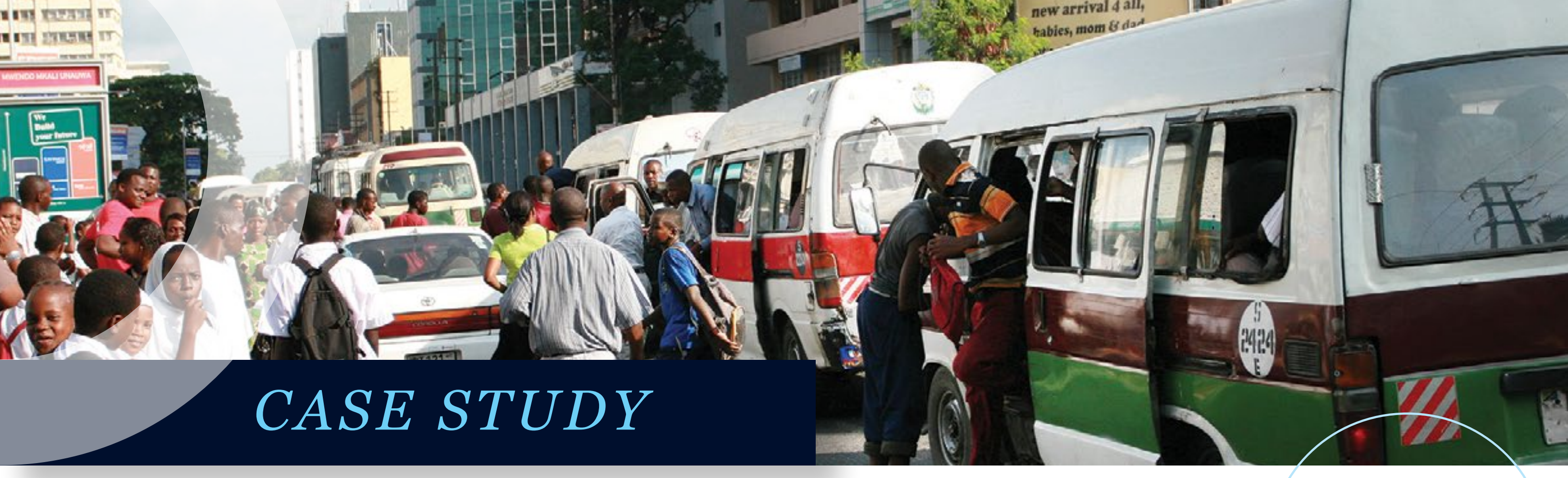
However, the team noted several limitations for their approach and made recommendations for future studies. First, the dataset was labeled according to 2014-15 imagery using a visual-interpretation method, which, by its nature, may be subject to idiosyncratic variation across individuals performing the manual classification. It is necessary to assure that across the dataset each example is labelled by multiple people and to account for the agreement between them. Second, the analysis is limited to India. Creating manually labeled ground-truth data is expensive and time consuming. However, crowd-sourcing platforms may allow researchers to scale – at low cost – the labeling method and to construct larger and more comprehensive ground-truth datasets. Third, the sampling method used in this study was designed to detect the boundaries between built-up areas and their periphery. The majority of the labeled examples were sampled from highly populated areas and from their adjacent, low-population environs. This approach may create a risk of false-positive detections when classifying distant or remote areas. It is therefore suggested that future projects taking this approach include in the training set examples from remote areas that are less populated.

New studies that take into account these limitations and successfully exploit emerging approaches such as crowd-sourced information will enable the measurement of urbanization with a high degree of geographic precision and in close to real time. This has the potential to transform how public policy is designed, helping planners to achieve successful urban growth and address many of the most persistent challenges of economic development.

Big Pixel, Case Study of India: University of California San Diego. Big Pixel Research Team: Ran Goldblatt with Wei You, Gordon Hanson and Amit K. Khandelwal

The study was published in Remote Sensing: Goldblatt, R.; You, W.; Hanson, G.; Khandelwal, A.K. Detecting the Boundaries of Urban Areas in India: A Dataset for Pixel-Based Image Classification in Google Earth Engine. Remote Sens. 2016, 8, 634.

Google Earth Engine: https://earthengine.google.com/



CASE STUDY

Big Data to Beat Congestion

Systems & Governance | Green Cities | Competitive Cities

NEAR-TERM

HIGHLIGHTS

- Growing cities are being challenged by the burden of large populations they were not built to accommodate. In response, they need to innovate to address problems of congestion and infrastructure improvement.
- Big data analytics can help cities better understand and manage traffic and infrastructure needs, even in circumstances where traditional data-collection methods cannot be applied.
- Successful approaches include congestion reduction models based on low-resolution traffic cameras; using cellphone data to understand people's intra-city travel needs, and analyzing GPS data from smartphones to assess traffic flow.

BEYOND TRAFFIC: POTENTIAL USES FOR DATA COLLECTION FROM CAMERAS AND SENSORS

- Track and manage city service delivery, such as waiting times in government offices, water and sewage systems or garbage collection
- Maintenance of law and order. Tracking crowd formation and criminal activity
- Disaster management and response

Practice Areas:
Systems and Governance, Green Cities, Competitive Cities

Countries Involved:
Kenya, Vietnam, Tanzania, Haiti, the Philippines

Data Types:
Low Resolution Cameras, Traffic Sensors

The United Nations estimates that the global urban population will grow by 2.5 billion by 2050, either through migration or the urbanization of rural areas. Ninety percent of this increase will occur in Asia and Africa, bringing the total global population living in urban areas to 66 percent[6]. Such population growth adds to the already significant challenges policymakers face to provide critical infrastructure and key services to urban populations. This is especially so in low- and middle-income countries, where urban development is frequently informal and incremental, and where cities struggle to keep up with infrastructure needs.

High among the challenges city officials face is congestion. Traffic congestion has negative impacts on economic growth and can exacerbate urban air pollution and greenhouse gas emissions. It hampers cities' resilience and competitiveness. In Nairobi, Kenya, for example, road congestion is estimated to cost around US$600,000 per day in lost productivity[7], and a considerable amount more if the impact of wasted physical resources and emissions are taken into account. Nairobi's infrastructure is decades old, planned for a city of around 350,000 inhabitants instead of the more than 3.4 million people who live there today. Improvements in public transport, roads and other critical infrastructure needed to support mobility in this expanded population require expensive investment. This needs to be carefully planned by drawing on diverse data sources. New construction and infrastructure development is typically a long process spanning several years, often involving the shutdown of existing transport facilities on which urban populations are critically dependent. But emerging techniques in big data analytics can track urban congestion and offer insights that pave the way for effective real-time mitigation measures.

High-quality insights from poor-quality data

Big data is already being used to tackle problems of urban mobility and transportation in many lower-income countries and rapidly growing cities. In Kenya, *Twende Twende* (Swahili for "let's go") is a platform developed by IBM's research center in Nairobi which uses predictive big data analytics to address congestion. At its core, Twende Twende takes images captured by existing low-cost cameras and applies network-flow algorithms to estimate traffic flow. The model overcomes challenges often associated with big data in low- and middle-income countries, such as a lack of data-collecting infrastructure, to deliver a cheap yet significant solution for congestion management.

The researchers parsed openly-available camera feeds from low-resolution traffic cameras, put in place by Access Kenya (a local internet provider), in collaboration with city

6. United Nations Report: World Urbanization Prospects (http://www.un.org/en/development/desa/news/population/world-urbanization-prospects-2014.html)
7. Source: http://asmarterplanet.com/blog/2013/10/27837.html

authorities[8]. They then overcame two major technical challenges that had prevented such a system from being developed previously: Firstly, the camera images were not of the quality which could be used by traditional computer vision algorithms to discern and count the number of individual cars. Secondly, the footage only covered about five percent of Nairobi's roads (as of 2013[9]).

IBM overcame these challenges by extracting unique data features and observations that had not been used in this context before. Instead of attempting to count the number of cars from the camera feeds and images, the researchers flipped the problem and tried to see how much of the road and background they could pick up with the presence of cars and other objects on it. This way, they were able to get a general sense of the percentage to which a road was congested – a data feature good enough to build a real-time model of traffic flow throughout the city.

To overcome the low percentage of roads covered by camera, the team took data from the five percent that were covered and used it to predict conditions on the "blind" 95 percent. They integrated data from their algorithm with empirical data from directly observed traffic in Nairobi's business districts, and built a predictive model of how traffic at key junctions would affect traffic on all the other connected roads. Estimates of traffic congestion from their predictive algorithm stood up well when compared to empirically collected traffic data.

Twende Twende is now being used to provide congestion warnings and route recommendations that users can access via SMS and mobile app-based map interfaces (through two of the country's largest mobile telecom providers – Airtel and Safaricom). This opens up a completely new avenue of crowdsourced data collection that would augment IBM's existing sources.

The data from the camera streams and inference of traffic patterns are also very useful for urban development and city officials, and other stakeholders wanting to improve transportation infrastructure. They can help policymakers prioritize development projects by objectively identifying critical infrastructure most in need of overhaul. They can also help in the effective planning of long-term projects, so that burdens to the population are minimized.

The power of open data

Central to Twende Twende's success has been the open availability of data from Access Kenya's traffic cameras, and a willingness to change the way that issues are typically addressed. Open data initiatives are promoting high-value data projects within and beyond government, providing businesses, academics, policy centers and other interested stakeholders with access to data that goes beyond revealing insights, but can also be leveraged to change existing power structures and decision-making processes.

Of great significance to lower-income countries is that Twende Twende overcomes the poor

8. http://traffic.accesskenya.com/
9. Source: http://www.forbes.com/sites/ehrlichfu/2015/03/03/fixing-traffic-congestion-in-kenya-twende-twende/

quality of data recorded by the low-resolution cameras by using smart algorithms and analytics techniques. Beyond traffic and congestion management, the platform's technology can be repurposed to create data streams valuable for numerous aspects of urban management – from monitoring, via closed-circuit TV footage, waiting times at offices that provide city services, to tracking garbage collection and sewage processing systems using sensors. Analytics tools are demonstrably effective in ensuring that numerous city facilities run smoothly. They provide efficient, low-cost solutions, boosting resilience and competitiveness in urban centers to help them keep up with growing populations. This potential means that development stakeholders must create the opportunities and infrastructure for data collection (even if minimal) and incentivize the provision of open access to it.

Next-generation congestion management

Twende Twende demonstrates how effectively big data can be gathered and used even within the constrained resources of growing cities. IBM's researchers are currently working to extend its capabilities to include data on accidents, weather conditions and roadworks to create a comprehensive view of human mobility[10].

More traditional avenues of big data collection are also being piloted in many developing countries and are proving useful in improving resilience and managing congestion. Da Nang is Vietnam's biggest seaport and fourth-largest city, with close to one million inhabitants. It is also the country's fastest growing metropolitan sprawl. Since 2013, as part of IBM's Smart Cities Challenge, Da Nang's traffic control center has had tools in place to predict and prevent congestion on the city's roads, and to better coordinate responses to situations caused by adverse weather or accidents. Data is aggregated from multiple streams, such as sensors embedded in roads and public transportation units. City officials analyze this to detect anomalies and control the flow of traffic. The system also provides the transport department with real-time information for its fleet of buses, allowing it to view details such as the location, speed and predicted journey times for each vehicle, and to respond to changing demand by adjusting services. Traffic data, along with sensors that monitor water levels on the flood-prone Han River, are facilitating early warning systems for disruptions to the city's transportation network, while also regulating activities at the port. Although Da Nang's traffic control system is in its nascent stages, the €37 million[11] project has already demonstrated the power of big data as a cost-effective approach to managing scarce infrastructure resources wisely.

Further innovative systems to manage congestion are emerging, with cellphone data and open-source software becoming increasingly important. In Dar es Salaam, Tanzania, World Bank economists Nancy Lozano Gracia and Talip Kilic are combining analysis of big data from sensor-embedded smartphones with face-to-face and phone interviews, to capture accurately and affordably the route,

10. http://www.research.ibm.com/articles/africa.shtml

11.Source: World Bank Background Document: BIG DATA AND URBAN MOBILITY (June 2014) (http://www.worldbank.org/content/dam/Worldbank/Feature%20Story/mena/Egypt/Egypt-Doc/Big-Data-and-Urban-Mobility-v2.pdf) purpose, travel mode and cost of every trip more than 500 individuals make within the city over a period of time. Reliable information about how individuals move around cities, and the constraints they face, enables policymakers to make informed, coordinated decisions on transport investments and land-use. Traditional methods for understanding individuals' travel patterns involve collecting travel diaries kept by respondents and supervised for completion by field staff. These methods are resource-intensive, subject to recall error and demanding on respondents. In response, the team sought to create a dataset highly informative about individual's travel patterns as a function of their socioeconomic background, the purpose of their travels and the associated costs.

After developing sensors and software that were installed in GPS-enabled smartphones, the team selected a random sub-sample of respondents to the World Bank's 2013-14 Measuring Living Standards Survey in Dar es Salaam. Respondents had already taken part in a face-to-face interview covering their socioeconomic background and travel patterns. Each was supplied with a smartphone able to collect and transmit the time and GPS location of individual movements at one-minute intervals for a one-month period. To encourage continued participation, respondents were told they could keep the phones after the study. Journey records from their phones were then validated via follow-up phone interviews every three days, covering the origin, destination, route, purpose and cost of each trip. Initial data analysis is now underway, focusing on understanding the key determinants for how people choose modes of transport.

The team aims to combine project data with cellphone call records to assess the feasibility of using lower-cost phone data for informing transport planning. Subsequent analysis will also examine how the travel patterns recorded compare to those based on traditional data sources, such as the Living Standards survey. The project's methodology could be applied to any city. With accurate understanding of people's travel needs and the challenges they face when moving around, city officials can make well-informed, coordinated decisions to optimize ease of movement and minimize congestion.

Cellphones are also at the center of a big data pilot in Haiti, which aims to carry out innovative analysis of call data records to provide valuable inputs for strengthening urban transportation and land use planning. These inputs can help planners unlock the economic potential of Haiti's cities and expand job opportunities. From the call data records, the pilot will identify key intra-city connectivity challenges, producing an employment accessibility analysis for the country's two biggest conurbations, Port-au-Prince and Cap-Haitien. Scenario analyses will be conducted to simulate potential interruptions in mobility between different parts of the cities, including from hazard and disaster events (such as flooding), and their impact on access to jobs. This will help identify key corridors that require resilient strategies to ensure people can reach employment in the event of a disaster.
The project will also include analysis of migration flows to and from urban areas to gauge access to opportunities across the country. Based on this analysis, the team will identify major bottlenecks and possible

interventions to improve city infrastructure, land use and coordination mechanisms, so as to promote inclusive employment. Access to jobs can be enhanced either by increasing travel speeds (through more or better roads and public transport) or by reducing distance to jobs. Based on the accessibility analysis, scenarios can be produced to test the potential for improvements in transportation and land use to enhance job access and social inclusion.

In the Philippines, the Open Traffic program leverages open-source software and innovative partnerships to substantially reduce the cost of traditional traffic data collection and analysis, while simultaneously improving the quality. The project team worked with the Cebu City Government to develop the first scalable, open-source platform of its kind for collecting, visualizing and analyzing traffic speed data derived from taxi drivers' smartphones. Using GPS data from an on-demand taxi service, Open Traffic successfully analyzed peak-hour congestion, travel time reliability and corridor vulnerability across 10 South-east Asian cities, and has prepared travel time analyses for select origin-destination pairs. This analysis would not previously have been possible without substantial time and resources. It will enable traffic management agencies to make better, evidence-based decisions about signal timings, public transport, road infrastructure, emergency traffic management and travel demand management.
These approaches all show that the next generation of congestion and travel management solutions will leapfrog the capital-intensive approaches of the past, enabling urban authorities and other stakeholders to make affordable, evidence-based planning decisions.

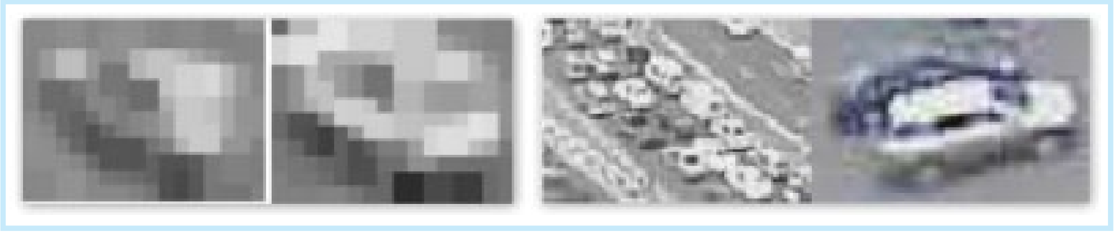


Figure 4: Low-resolution traffic cameras show images of cars as unclear (left) or overlapping (right). Twende Twende's innovative approach avoids these problems by analyzing the images for roadspace, rather than for vehicles.

Source: IBM Research/Forbes Magazine

CASE STUDY

People Power: Crowdsourcing Data to Track Crime

Resilient Cities | Inclusive Cities | Systems & Governance

DEVELOPMENT (3-5 years)

HIGHLIGHTS

- Reducing crime and violence is a priority for public policy stakeholders looking to support inclusive, resilient and competitive urban environments.
- The proliferation of technology is opening new avenues for data collection through crowdsourcing, for tracking, analyzing and responding effectively to criminal activity in urban areas. Crowdsourced data can also be leveraged to predict and proactively prevent crime.
- Open-source crowd-mapping platforms can be used to monitor a wide range of factors, including election fairness, corruption, natural disasters, and the supply of utilities and urban services.

CROWDSOURCED DATA: POLICY IMPLICATIONS

- Dynamic feedback to evaluate the effectiveness of measures and to allow programs to be refined
- Increased transparency and government accountability
- Efficient information sharing for targeted management in crisis situations
- Predictive insights into future amenity and infrastructure needs of urban populations, through crowdsourcing combined with other data streams

Practice Areas:
Resilient Cities, Inclusive Cities, Systems and Governance

Countries Involved:
Colombia, Kenya, Syria

Data Types:
Crowdsourcing, digital listening, social media

Crime and violence are now a key development issue in many lower-income countries worldwide. Beyond the trauma and suffering of individual victims, they carry staggering economic costs at both local and national levels. Accounting for expenditures on citizen security, law enforcement and health care, experts estimate that crime costs close to 8 percent of Gross Domestic Product (GDP) in some Central American countries. Crime and violence also undermine economic growth, not just through the loss of victims' wages and labor, but by polluting the investment climate and diverting scarce government resources from supporting economic activity to strengthening law enforcement. Some estimates suggest that a 10 percent reduction in violence in Central American countries with the highest murder rates could boost annual economic growth per capita by as much as one percent.[12]

Reducing crime and violence is a priority for stakeholders looking to support inclusive, resilient and competitive urban environments. One of the largest problems with tackling violent criminal activity in low- and middle-income countries is the absence of conventional frameworks for the lateral reporting of crime. This results in an environment in which government authorities and policymakers are unable to identify and focus targeted law enforcement efforts on problematic areas, while crime victims and the general population lack trust in the criminal justice system.

Traditionally, big data has not been applied to the tracking and management of crime among lower-income cities owing to the lack of data collection and analysis mechanisms. Poor law enforcement infrastructures reflect crime statistics that are rarely collected and organized, or largely unreliable or incomplete. However, emerging technology has the potential to open a completely new avenue of data collection for tracking criminal activity: Crowdsourcing.

Mapping crime through crowdsourcing

In Colombia's capital, *Pilas Bogota* (which loosely translated means "get sharp, Bogota") is the first crime map of its kind that draws on crowdsourced information from victims and witnesses of crime. The platform was developed in 2011 by a "Hacks Hacker" chapter in Colombia (a collective of technologists and app developers), in collaboration with the International Center for Journalists. It collects the time, date and location at which incidents occur from victims, witnesses and other citizens, and plots the collected information onto an interactive dynamic map. This is actively administered and monitored by journalists at *El Tiempo*, a local news media outlet. They trawl the information for tips and trends to produce more thorough and in-depth reporting about crime in the

12. Source: World Bank Report - Crime and Violence in Central America: A Development Challenge (2011) (http://siteresources.worldbank.org/INTLAC/Resources/FINAL_VOLUME_I_ENGLISH_CrimeAndViolence.pdf)

city. Citizens can also report incidents by sharing posts and images or video using social media, with geo-located tags and hashtags. These are picked up by Pilas Bogota's digital listening components or by mobile SMS – through which people can respond to short surveys about events close to them.

Pilas Bogota rests on the open-source Ushahidi Crowdmap platform, initially developed to map reports of post-election violence in Kenya in 2008, using information submitted via the web and mobile phones. Since its initial deployment, the platform has been repurposed for data collection and monitoring many crisis events, such as water shortages, earthquakes and natural disasters[13], as well as routine monitoring of high-crime areas or during elections.

Pilas Bogota has largely been used so far for collecting citizen reports of crime for journalistic purposes. However, the crowdsourcing approach to visualizing incidents in real-time is also useful for policymakers and city authorities, who have in the past relied solely on lagging police statistics and official reporting for the tracking and management of criminal activity. Even in cases where this information is reliable and complete, it can only provide a partial picture of the ongoing crime situation on the ground, as only a fraction of crime that occurs is reported or acted on by law enforcement agencies. As a result, non-profit and citizen-run organizations are embracing crowdsourced data not only for reporting crime, but also to augment the management of other city operations, including transport, services and corruption monitoring. Two additional crowdsourcing platforms have been launched: *Mi Bogota Verde*, to map Bogota's ongoing problems with garbage accumulation, and *Monitor de Corrupción*, to monitor corruption.

Combining datasets for added strength

Crowdsourcing gives city authorities the advantage of being able to report and analyze data in real time. This enables citizens, civil society, law enforcement agencies and humanitarian response agencies to respond quickly. During the Kenyan general election of 2013, a Ushahidi-based tool named *Uchaguzi* (Swahili for "election") was deployed. The process involved partnerships with civil society organizations and citizen groups, and used the same techniques as Pilas Bogota: Web and online reporting, coupled with social media and mobile SMS input. The platform crowdsourced data around key events and possible criminal activity during the election, followed by verification and appropriate escalation. This facilitated real-time intervention by the appropriate authorities.

Based on the success of this deployment, the Ushahidi platform launched "The Resilience Network Initiative". This provides city governments with online tools and the capacity to connect more closely with their citizens, collecting and sharing data and information, and allowing various stakeholders to input for improved municipal decision-making.

As with other types of data, crowdsourcing data becomes significantly more valuable from

13. http://www.ushahidi.com/mission/

a visualization and analysis perspective when combined with other datasets. The same open-source Ushahidi-based platform was deployed in Syria in 2012 to compliment an open-source web and social media tracking platform that mines thousands of online sources for evidence of human rights violations, killings, torture and detainment. The crowdsourcing tool, called Syria Tracker, coupled the open-source data with crowd-sourced human intelligence, such as field-based eye-witness reports shared via webform, email, Twitter, Facebook, YouTube and voicemail[14].

Using this approach, the Syria Tracker team and its relatively small group of volunteers have been able to verify almost 90 percent of the documented killings mapped on the platform, thanks to video or photographic evidence. They have also been able to name around 88 percent of those reported killed by Syrian forces since the uprising began. Depending on the levels of violence, the turnaround time for a report to be mapped on Syria Tracker is 1-3 days. The team produces weekly situation reports based on the data collected, along with detailed graphical analysis. Files providing a more precisely geo-tagged tally of deaths per location are made available regularly and can be uploaded and viewed using Google Earth.

This approach could easily be applied to other contexts and issues, such as petty crime in a city, women commuting through unsafe urban areas, environmental hazards, corruption and challenges to urban infrastructure facing quickly growing populations.

14. http://irevolution.net/2012/03/25/crisis-mapping-syria/, Retrieved 6/29/2015

Crowdsourcing for predictive analytics

Crowdsourced data collected over longer periods of time also contains significant potential for the measurement and evaluation of steps that city authorities and policymakers are taking towards the mitigation of crime and violence. It is often difficult to ascertain which policies work best to ensure safe urban environments, especially if official reporting and police statistics are lagged and are the only data sources available. Policymakers can benefit from the dynamic feedback that crowdsourced data streams offer in order to refine programs and enforcement efforts as they take place. This kind of dynamic interaction provides predictive insights, helping to shape operations and optimize use of government resources.

As more and more people use smartphones and social media, urban law enforcement officials in high-income countries are increasingly using crowdsourced data for gathering information and evidence in crisis situations, such as terrorist attacks, civil unrest or natural disasters. A wealth of geo-coded and time-stamped data is emerging for authorities to use in their efforts to piece together ongoing events and prepare appropriate responses. Some law enforcement departments have started using crowdsourcing technology to harness the public in trying to find missing property and capture petty criminals in their communities.

When coupled with other disparate datasets, crowdsourced data can go beyond simple descriptive analysis, offering the predictive and prescriptive analysis needed to proactively deter urban crime. Numerous studies show that if treated like a contagion, crime could be predicted from historical events[15]. In California, the Santa Cruz Police Department is using advanced analytical techniques to predict property crimes such as home burglaries and car thefts, and deploying officers to suspected locations in advance. Based on models for predicting aftershocks from earthquakes, their approach analyzes and detects patterns in past crime data, to generate projections about which areas and times are at highest risk of future crimes. The projections are recalibrated daily, as new crimes occur and updated data is fed into the program[16].

All these approaches highlight the importance of publicly available, open-source mapping platforms. These can harness the knowledge and experiences of ordinary people in the quest to develop safe cities.

Current Projects on the Ushahidi Platform:
Pilas Bogota: Citizen and Journalist crime tracking utility
Monitor de Corrupción: Crowdsourced corruption monitoring
Uchaguzi: Election monitoring mechanism
Syria Tracker: Combined crowdsourced information with digital listening for crisis monitoring

Figure 5: An interactive page on Pilas Bogota, with news reports and collection prompt for crowdsourced information (bottom right)

15. http://www.nber.org/papers/w12409
16. http://www.nytimes.com/2011/08/16/us/16police.html?_r=0

CONCLUSION: BIG DATA FOR OPTIMAL CITIES OF THE FUTURE

DEVELOPMENT (up to 10 years)

Informed, considered and coordinated urban planning is needed to build sustainable, resilient, equitable and livable cities. To plan effectively, stakeholders need detailed descriptive and predictive information on both macro and micro levels – from large-scale trends such as the nature and pace of urbanization, to the needs of individual communities and citizens. The cities of the future must raise real incomes of the poorest people, be resilient to shocks and stresses, protect environmental resources and actively improve the lives of their residents.

Big data can help bring about step changes in all these areas, from national to local government levels. Initiatives like those profiled in this brief are just the beginning – both in terms of their own lifecycle and impact, and in terms of the variety of ways in which big data analytics will help cities become better places. On the horizon are many more big data approaches, which look set to help streamline urban planning and management in the coming decade. Some of the most exciting include:

Collecting taxes through geo-spatial data
Stakeholders: Local authorities, policymakers
Data Types: Geo-spatial information (satellite), drones

The availability of Geo-spatial Information Systems (GIS) such as high-resolution satellite imagery and drones has resulted in an increased use of these data sources in urban planning and development. So far, this has mainly been at national or macro policy level. However, integration of insights derived from GIS data into local urban planning and management offers great potential. A recent feasibility study into integrating GIS from the bottom up into Ghana's Land Administration Project evaluated local administration technological and organizational preparedness for GIS use. It also laid out a policy roadmap for GIS adoption from national databases and systems to the local administrative levels, for a pilot project to improve property tax collection[16]. On a wider scale, the World Bank has funded a local government revenue collection system using a GIS platform in 16 municipalities in Tanzania, under its Urban Local Government Strengthening Program[17]. This management tool supports local government tax reporting, revenue collection, operations and maintenance, urban planning, licensing and land management systems. It means analysis of big data from GIS systems can deliver valuable insights beyond the central government level, direct to city authorities.

16. Study carried out by researchers working for the annual paper competition for advanced graduate studies on issues relating to urban poverty, co-sponsored by USAID, the Wilson Center, the World Bank, the International Housing Coalition and the Cities Alliance.
17. Source: http://www.worldbank.org/projects/P118152/tanzania-second-local-government-support-project?lang=en

Mining mobile money transactions for insight
Stakeholders: Individuals, businesses, policymakers
Data Types: Mobile

Cellphones have become ubiquitous and are used increasingly by people in the world's poorest countries. In 2012, there were 5.9 billion active mobile connections globally, forecast to increase to 7.6 billion in 2017. Much of this increase will come from cellphone penetration in Africa and Asia. In Africa, services like *mPesa*, offering mobile wallet and digital transactions, are increasingly important, with many sections of the urban population choosing them for everyday financial transactions. Figures from Kenya's Central Bank showed that for the year ending November 2014, the value of mobile money transactions rose 26 percent, to 1.94 trillion Kenyan shillings (KES), while card transactions fell 18 percent to KES1.1 trillion[18.] As mobile money transactions are possible between individuals, they are fast replacing cash as the preferred medium of exchange between people. Today, many development practitioners are either hesitant to use, or prohibited from using, cellphone transaction data, owing to the lack of policy frameworks and privacy protection mechanisms. As the value of using this data becomes more apparent (revealing pockets of poverty through low transaction amounts, for example), such restrictions will be removed and adequate data-sharing frameworks will be put in place. These will protect individuals' privacy, while enabling the wealth of big data

19. Source: Kenyan Central Bank payment systems data (http://www.thepaypers.com/mobile-payments/kenya-mobile-money-hits-usd-21-bln-as-transactions-surge/758093-16)25

generated by cellphones and mobile money to be harnessed for development.

Understanding how infrastructure affects crime

Stakeholders: Local authorities, individuals, businesses, policymakers
Data Types: Geo-spatial

Latin America is highly urbanized, with above-average crime rates. Its cities are typically unplanned, with high socio-economic inequality, yet the association between crime and infrastructure has not been clearly defined or quantified. Colombia's capital, Bogota, collects considerable geo-coded data on urban infrastructure and has reliable geo-coded information on population and crime. Drawing on this rich data, a World Bank project quantified the occurrence of crimes in relation to specific characteristics in the built environment, through a technique called risk terrain modeling. This uses an algorithm that identifies relationships between different layers of data and correlates them with crime, using models which are then linked to places on a digitized map. The approach combines separate layers of map (one per risk) to produce maps showing the intensity of all risk factors at every location throughout a landscape – the 'risk terrain'. This allowed the team to identify locations near bus stations, public hospitals, schools and drugstores as being associated with assault and homicide. The modeling also revealed peak times of day for crime, and predicted areas of the city more likely to experience future crime. Combined with local stakeholder interpretations, risk terrain mapping can reliably suggest action to reduce crime associated with particular environmental factors. The methods are widely applicable in other locations and for other crimes.

A rich learning process

The potential for big data is huge, but it is not a magic bullet. Innovative paths inevitably involve hurdles, reveal useful lessons and require perseverance. Big data demands that users capture, prepare and store data meticulously – and plan enough time to do so. For the Twende Twende congestion management system in Kenya, a critical first step was the availability of data. To enable big data analytics, it is essential that city authorities, policymakers, non-governmental organizations and other stakeholders promote the public availability of useful data from a wide range of sources (with policies to support its use).

Successful big data solutions often involve approaching existing situations from new angles or combining previously unrelated data sources – such as using EO data, captured for other purposes, to assess urban characteristics and growth. Despite the central role of computational power, the human element also remains vital to the success of these projects. Approaches such as the Ushahidi crowdsourcing platform are reliant on human input. Big data analysis can often be enhanced by traditional research techniques, such as socio-economic surveys. There is also need to invest in partnerships, or to combine human and computational power for optimum results. Big data analytics is a team sport: Effective collaboration between data experts, technologists and business sector specialists is crucial.

All these novel data approaches must be tested, validated and adapted for mainstream use – but the potential rewards of big data make such effort worthwhile. As the case studies in this brief show, big data analytics can improve development effectiveness in cities and help initiatives within and beyond the World Bank achieve results through improved evidence, efficiency, awareness, understanding and forecasting. Ultimately, big data initiatives can be a powerful accelerator for ending poverty and boosting shared prosperity.

REFERENCES

Using Geospatial Data to Track Changes in Urbanization

World Bank group. (2013). Building African Cities that Work: A study on the Spatial Development of African Cities (Concept Note). World Bank Group.

World Bank Group. (2015). East Asia's Changing Urban Landscape: Measuring a Decade of Spatial Growth. World Bank Group.

Big Data to Beat Congestion

Bills, T., Bryant, R., & Bryant, A. W. (2014, October). Towards a frugal framework for monitoring road quality. In Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on (pp. 3022-3027). IEEE.

Ehrlich, T., & Fu, E. (2015, March 3). Fixing Traffic Congestion In Kenya: Twende Twende. Retrieved from Forbes: http://www.forbes.com/sites/ehrlichfu/2015/03/03/fixing-traffic-congestion-in-kenya-twende-twende/

IBM Research. (n.d.). IBM Resarch - Africa: Developing solutions in Africa, for Africa and the world. Retrieved from IBM Research Articles.

Kinai, A., Bryant, R. E., Walcott-Bryant, A., Mibuari, E., Weldemariam, K., & Stewart, O. (2014, June). Twende-twende: a mobile application for traffic congestion awareness and routing. In Proceedings of the 1st International Conference on Mobile Software Engineering and Systems (pp. 93-98). ACM.

Larsen, L. (2013, Nov 5). IBM Nairobi Lab's First Offering is a Traffic-Dodging Mobile App. Retrieved from IEEE Spectrum: http://spectrum.ieee.org/tech-talk/computing/software/ibm-nairobi-labs-first-offering-is-a-trafficdodging-mobile-app

Witchalls, C. (2013). Text messages tell drivers when there's a jam ahead. New Scientist, 220(2944), 23.

World Bank Group. (2014). Big Data and Urban Mobility. Retrieved from http://www.worldbank.org/content/dam/Worldbank/Feature%20Story/mena/Egypt/Egypt-Doc/Big-Data-and-Urban-Mobility-v2.pdf

People Power: Crowdsourcing Data to Track Crime

Kling, Jeffrey R. J. L. (2005). Is Crime Contagious? CEPS Working Paper No. 117.

Knight International Journalism Fellowships. (2012). Colombia: Use Crowd Sourcing Technology to Track Crime and Corruption. Retrieved from International Center for Journalists (ICFJ): Our Work: http://www.icfj.org/our-work/colombia-use-crowd-sourcing-technology-track-crime-corruption

Lovler, R. (2012, Oct 17). Colombian Media Rely on Mapping to Track Crime, Corruption, Environmental Issues. Retrieved from International Center for Journalists (ICFJ) - Blogs: http://www.icfj.org/blogs/colombian-media-rely-mapping-track-crime-corruption-environmental-issues

Lovler, R. (2012, March 26). Mapping Crime and Corruption in Colombia: Knowledge is Power, Thanks to New Digital Technology. Retrieved from International Center for Journalists: Blogs: http://www.icfj.org/blogs/mapping-crime-and-corruption-colombia-knowledge-power-thanks-new-digital-technology

Meier, P. (2012, March 25). Crisis Mapping Syria: Automated Data Mining and Crowdsourced Human Intelligence. Retrieved from iRevolutions: From Innovations to Revolutions: http://irevolution.net/2012/03/25/crisis-mapping-syria/

Meier, P. (2012). Crisis mapping in action: How open source software and global volunteer networks are changing the world, one map at a time. Journal of Map & Geography Libraries, 8(2), 89-100.

Okolloh, O. (2009). Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. Participatory learning and action, 59(1), 65-70.

VanCalcar, J. E. (2006). Collection and representation of GIS data to aid household water treatment and safe storage technology implementation in the Northern Region of Ghana (Doctoral dissertation, Massachusetts Institute of Technology).

ZULUAGA, Á. B. (2012, Oct 16). *Sea un reportero de la seguridad*. Retrieved from *El Tiempo:* Archivo: http://www.eltiempo.com/archivo/documento/CMS-12306723

Conclusion: Big Data for Optimal Cities of the Future

Cartesian; Bill & Melinda Gates Foundation. (2014). Using Mobile Data for Development. Cartesian. Retrieved from http://www.cartesian.com/wp_content/upload/Using-Mobile-Data-for-Development.pdf

Diko, S. K., & Akrof, A. A. (2011). Optimizing Property Rate Returns for Urban Development in Ghana, Using Geographic Information Systems. Innovation in Urban Development: Incremental Housing, Big Data and Gender (Wilson Center and USAID Report), 146-165.

World Bank Group. (2014, May 30). Projects & Operations. Retrieved July 28, 2015, from http://www.worldbank.org/projects/P118152/tanzania-second-local-government-support-project?lang=en

ACKNOWLEDGEMENTS

This solutions brief is one of several sector focused big data knowledge products delivered through the "Innovations in Big Data Analytics" program which resides in the Global Operations and Knowledge Management Unit (GOKMU). It has been prepared jointly in cooperation with the Social, Urban, Rural and Resilience Global Practice and Booz Allen Hamilton. Trevor Monroe (GOKMO), Andrew Whitby (DECDG), and Luda Bujoreanu (GTIDR) coordinated the publication. A number of WBG staff provided valuable contributions to make this publication possible, including peer reviewers Ellen Hamilton, Nancy Lozano Gracia, Katherine Kelm, and Kai Kaiser.

The *Innovations in Big Data Analytics* program works to accelerate the effective use of big data analytics across the World Bank, and to position the World Bank as a leader in the big data for development community.

For additional information about this brief or to find out more about the program, please visit http://bigdata (WBG intranet) or contact Adarsh Desai (adesai@worldbank.org) or Trevor Monroe (tmonroe@worldbank.org).

WORLD BANK GROUP