



COMMITTED TO
IMPROVING THE STATE
OF THE WORLD

Big Data, Big Impact: New Possibilities for International Development

Executive Summary

A flood of data is created every day by the interactions of billions of people using computers, GPS devices, cell phones, and medical devices. Many of these interactions occur through the use of mobile devices being used by people in the developing world, people whose needs and habits have been poorly understood until now. Researchers and policymakers are beginning to realise the potential for channelling these torrents of data into actionable information that can be used to identify needs, provide services, and predict and prevent crises for the benefit of low-income populations. Concerted action is needed by governments, development organisations, and companies to ensure that this data helps the individuals and communities who create it.

Special Thanks

The World Economic Forum acknowledges the work of Vital Wave Consulting in assembling this briefing and the extensive research conducted by partners such as the Boston Consulting Group, the UN Global Pulse, the Global Viral Forecasting Initiative, Ushahidi, the World Bank, Vodafone, and the World Economic Forum's ICT Global Agenda Council on the transformative potential for harnessing big data.

Big Data, Big Impact: New Possibilities for International Development

By analysing patterns from mobile phone usage, a team of researchers in San Francisco is able to predict the magnitude of a disease outbreak half way around the world. Similarly, an aid agency sees early warning signs of a drought condition in a remote Sub-Saharan region, allowing the agency to get a head start on mobilising its resources and save many more lives.

Much attention is paid to the vital services that mobile phone technology has brought to billions of people in the developing world. But now many policy-makers, corporate leaders and development experts are realising the potential applications, like the examples above, for the enormous amounts of data created *by and about* the individuals who use these services.

Sources such as online or mobile financial transactions, social media traffic, and GPS coordinates now generate over 2.5 quintillion bytes of so-called 'big data' every dayⁱ. And the growth of mobile data traffic from subscribers in emerging markets is expected to exceed 100% annually through 2015ⁱⁱ.

The data emanating from mobile phones holds particular promise, in part because for many low-income people it is their only form of interactive technology, but it is also easier to link mobile-generated data to individuals. This data can paint a picture about the needs and behaviour of individual users rather than simply the population as a whole.

Building user-centric solutions offers compelling possibilities for providing better access to services in health, education, financial services, and agriculture for people living in poverty.

Financial Services

Data gleaned from mobile money services can provide deep insight into spending and saving habits across sectors and regions. Digital payment histories can allow individuals to build credit histories, making them candidates for loans and other credit-based financial services.

Education

Data derived from the use of mobile value-added services can be used to improve public-sector understanding of educational needs and knowledge gaps, allowing more targeted and timely initiatives to disseminate critical information.

Health

Data collected through mobile devices, whether captured by health workers, submitted by individuals, or analysed in the form of data exhaust, can be a crucial tool in understanding population health trends or stopping outbreaks (see box on page 5). When collected in the context of individual electronic health records, this data not only improves continuity of care for the individual, but it can be used to create massive datasets with which treatments and outcomes can be compared in an efficient and cost effective manner.

Agriculture

Mobile payments for agricultural products, input purchases and subsidies may help governments better predict food production trends and incentives. This knowledge can be used to ensure the availability of proper crop storage, reduce waste and spoilage, and provide better information about what types of financial services are needed by farmers. Mobile use patterns may also help governments and development organisations identify regions in distress so that targeted assistance can be directed to them. Early detection can help prevent families from leaving their land and further decreasing agricultural production.

Data through the Mobile Financial Services Lens

A look at the 'pillars' from The World Economic Forum's Mobile Financial Services Development Report offers insights into how the requirements of mobile financial services development coincide with prerequisites for a thriving data commons.

Regulatory Proportionality: Finding appropriate uses for mobile-generated data will require regulation similar to that needed for mobile financial services. In both situations, regulation must keep pace with new technology and protect consumers without stifling innovation or deterring uptake. The development of sensible data standards could increase uptake of both mobile financial services and individual data security.

Consumer Protection: As with mobile financial services, proper regulation and data ownership processes must be put in place to prevent the theft or misuse of sensitive information.

Market Competitiveness: In the long term, adequate competition is essential to ensure a wider range of affordable services and interoperability. However, private-sector companies should be encouraged to allow access to non-sensitive data that can benefit populations and deepen their own understanding of individual behaviour. Such cooperation may also help telecom operators realise that creating interoperable mobile money systems can benefit them over the long term.

Market Catalysts: For both the data commons and mobile money, government can serve as a catalyst to ensure legitimacy. This will require open and transparent governance, as the idea of government access to an individual's financial information could discourage uptake of mobile financial services.

End User Empowerment & Access: Individuals must have a moderate degree of financial literacy, affordable access to a mobile device, and a mobile network connection, in addition to control over their own information.

Distribution and Agent Network: Analysing transactional data could determine where there is demand for additional mobile money agents.

Likewise, utilising the data created by mobile phone use can improve our understanding of vulnerable populations, and can quicken governments' response to the emergence of new trends. Actors in the public, private, and development sectors are beginning to recognise the mutual benefits of creating and maintaining a 'data commons' in which this information benefits society as a whole while protecting individual security and privacy. But a more concerted effort is required to make this vision a reality.

Understanding the Dynamics of the Data Ecosystem

To turn mobile-generated data into an economic development tool, a number of ecosystem elements must be in place. For those individuals who generate the data, mechanisms must be developed to ensure adequate user privacy and security. At the same time, business models must be created to provide the appropriate incentives for private-sector actors to share and use data for the benefit of society. Such models already exist in the Internet environment. Companies in search and social networking profit from products they offer at no charge to end users because the usage data these products generate is valuable to other ecosystem actors. Similar models could be created in the mobile data sphere, and the data generated through them could maximise the impact of scarce public sector resources by indicating where resources are most needed.

A look at the various types of data and actors in the data ecosystem illustrates the roles and incentives at work. The private sector maintains vast troves of transactional data, much of which is 'data exhaust', or data created as a by-product of other transactions. With the use of mobile phones, much of this data can be associated with individuals and their locations. The public sector in most countries also maintains enormous datasets in the form of census data, health indicators, and tax and expenditure information.

The Internet and mobile revolution have added yet another source: data contributed by

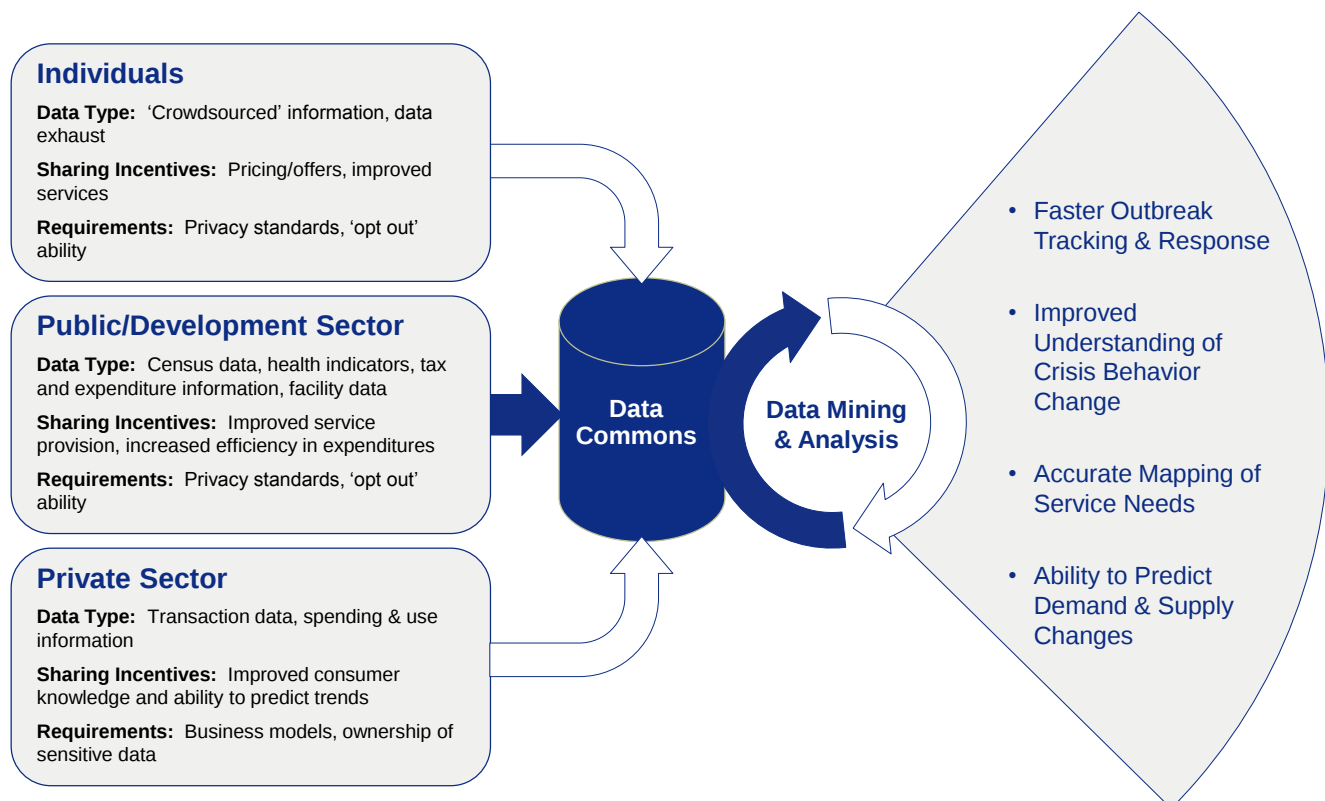
individuals voluntarily or through crowdsourcing. NGOs like Ushahidi are already using crowdsourcing to obtain, verify and disseminate real-time information about natural disasters and election monitoring, and Ushahidi is developing ways to filter and use the huge amounts of information being created with applications such as SwiftRiverⁱⁱⁱ. With the overwhelming majority of the world now having access to a mobile phone^{iv}, crowdsourcing allows individuals to contribute to the information gathering process, making it more democratic and transparent. The graphic below illustrates the various data types, incentives, and requirements of actors in this new data ecosystem.

Closing the Information Gap: Identifying the Returns from Better Data Use

Already, a number of organisations in the public and development sectors have embraced the vision of a data ecosystem in which information captured from these varied sources is used for the benefit of global populations. Global Pulse is a UN initiative aimed at bringing together expertise from the public, private, development, and academic sectors to develop approaches

for harnessing data for policy and action. Its director, Robert Kirkpatrick, says that data collected through mobile device usage can spur effective action in two primary ways: by reducing the time lag between the start of a trend and when governments and other authorities are able to respond to them, and by reducing the knowledge gap about how people respond to these trends.

Kirkpatrick cites Dr. Nathan Eagle's research^v showing that when mobile operators see airtime top-off amounts shrinking in a certain region, it tends to indicate a loss of income in that population. Such information might indicate increased economic distress before that data shows up in official indicators. Meanwhile, Global Pulse's own research into food related conversations on Twitter has shown very strong correlations with food price inflation^{vi}. "This information comes from two brand new sources: what people are doing and what they are saying," says Kirkpatrick. "As a government or aid agency, you might know that food prices are rising or rains aren't coming, but what if you could see where and how people are already



changing their behaviour and prioritise where you put resources in response?”

Public health offers one of the most compelling areas where the analysis of mobile and Internet data could lead to huge public gains. The San Francisco-based Global Viral Forecasting Initiative (GVFI) uses advanced data analysis on information mined from the Internet to identify comprehensively the locations, sources and drivers of local outbreaks before they become global epidemics. GVFI's Chief Innovation Officer, Lucky Gunasekara, says this technique can successfully predict outbreaks up to a week ahead of global bodies such as the World Health Organisation that rely on traditional techniques and indicators.

Employing new data collection and analysis methods could be a less costly, more efficient method of developing market intelligence for large organisations like the World Bank. The Bank already spends millions of dollars each year on statistical analysis of the needs of the poor^{vii}. Smarter data collection and analysis could free resources for use in economic development efforts^{viii}.

In a time of constrained government resources and reduced foreign aid, the insight produced by mining mobile data offers the possibility of preventing crises and targeting services to the populations that need them most. Yet there are serious challenges that need to be addressed before the pieces of the puzzle fall into place.

Obstacles on the Path to the Data Commons

Ecosystem actors, like those described above, have much to gain from the creation of an open data commons. Yet the sharing of such data especially that tied to individuals raises legitimate concerns that must be addressed to achieve this cross-sector collaboration.

Privacy and security: As ecosystem players look to use mobile-generated data, they face concerns about violating user trust, rights of expression, and confidentiality. Privacy and security concerns must be addressed before

firms, governments, and individuals can be

Mobilising Data to Deal with an Epidemic

In the wake of Haiti's devastating 2010 earthquake, researchers at the Karolinska Institute and Columbia University demonstrated that mobile data patterns could be used to understand the movement of refugees and the consequent health risks posed by these movements. Researchers from the two organisations obtained data on the outflow of people from Port-au-Prince following the earthquake by tracking the movement of nearly two million SIM cards in the country. They were able to accurately analyse the destination of over 600,000 people displaced from Port-au-Prince, and they made this information available to government and humanitarian organisations dealing with the crisis. Later that year, a cholera outbreak struck the country and the same team used mobile data to track the movement of people from affected zones. Aid organisations used this data to prepare for new outbreaks. The example from Haiti demonstrates how mobile data analysis could revolutionise disaster and emergency responses.

convinced to share data more openly.

Data personalisation: When individuals have multiple SIM cards, it is impossible to aggregate data from each SIM back to the same individual. This data is most useful if it can be attached to demographic indicators, which allow the data to tell a story about the habits of a segment of the population. Improved methods of tying subscriptions to demographic information are needed to ensure data generated by mobile devices is as individualised as possible.

Data sharing incentives: Individuals, fearing security and privacy concerns, often resist sharing personal data. In addition, many private-

sector firms do not see an incentive to share data they regard as proprietary. Governments often cannot force contractors to share data collected in the execution of public contracts or make all government data available for use by academia, development organisations, and companies. All players must see material benefits and incentives in data sharing that outweigh the risks.

Human Capital: Accurate and actionable data mining and analysis requires considerable technical skill, and data scientists are both in short supply and expensive to employ. GVFI's Gunasekara notes that even many large corporations do not have access to the type of expertise they need to develop novel data mining techniques. "Most of these people want to start their own companies, not work for someone else," he says. Maximising the contribution of human capital requires incentives for these individuals to use their talents for the public good along with long-term efforts to grow the talent pool.

Overcoming the Obstacles: Novel Approaches

A number of organisations are already working to overcome the challenges and create the incentive structures needed for cross-sector cooperation. Global Pulse is creating a network of Pulse Labs that bring together experts in government, academia, the development sector, and private companies to pioneer new approaches to using data for development challenges.

The organisation is also now actively engaging with partners around what Robert Kirkpatrick calls 'data philanthropy', where corporations are encouraged to share anonymised data for use by the public sector to protect vulnerable populations. These companies are driven partly by a recognition that more effective policy action will lead to greater resilience from economic shocks, and therefore translate into better business continuity. Athletic apparel company Nike has demonstrated an approach to corporate data sharing through its GreenXchange patent-sharing system. Nike is among the first corporations to explore opening up data publicly, and plans to share data on the sustainability of its operations.

In the area of individual incentives, Jana, a Boston-based start-up, conducts market research for global organisations in over 50 countries. The company uses SMS to survey emerging market customers in exchange for airtime, creating a financial incentive for consumers to overcome their concerns about sharing personal data.

Nathan Eagle, Jana's founder, notes that the data created is useful not only to marketing organisations in private-sector companies, but has extensive development uses as well. He cites a mobile data analysis effort in the huge slum of Kibera, outside Nairobi, that was used to map population change and direct latrine and water pipe building efforts for the benefit of the slums residents.

Government as Data Catalyst

Several forward-thinking governments in the developing world are demonstrating how government can catalyse the development of this ecosystem through the opening of its own datasets and the active management of their dissemination and use.

In July 2011, Kenya launched its new Open Data Portal, which includes a full digital edition of the 2009 census, 12 years of detailed government expenditure data, government household income surveys, and the location of schools and health facilities. The portal provides unlimited data access on the web and through mobile phones to researchers, web and software developers, journalists, students, civil society and the general public. Civic organisations, mobile application developers, and media groups are already using the data to improve understanding of population patterns, increase the transparency of governments, and map public services.

The World Bank has provided support for the initiative, but Chris Finch at the World Bank notes that the Kenyan authorities, with support at the highest levels of government, drove the initiative forward. In part, the Kenyan government responded to the country's growing information technology sector and the new constitution's guarantees of access to information. Finch notes that the government moved forward with the Open Data Portal before the legislative, policy and legal

framework, including protections for data reuse, were fully in place, given momentum around Constitutional guarantees for openness, transparency, and participation. Policy frameworks

Government Catalyst

- Enact appropriate legislation protecting end users without stifling innovation
- Open data to the public (free or for purchase) in a way that allows for innovation without infringing on citizen's privacy
- Encourage the development of appropriate technological infrastructure and training of individuals capable of analyzing big data

Public-Private Collaboration

- Telecoms and governments must work together to find a way to track mobile information back to an individual, rather than a SIM
- Government or Multi-lateral funded initiatives using data generated from mobile for development or government planning purposes (e.g., health, agriculture, education)

for such protections are now being backfilled. Finch sees the role of government as setting the legal frameworks governing data privacy and security, and also in developing systems that allow various agencies and ministries to continually update the data they make available. The development community can encourage this behaviour by supporting progressive governments such as Kenya's and linking them to the technical and financial resources they need. Unfortunately, in many countries, governments are frequently seen as part of the challenge to establish a productive data commons. Kenya's example demonstrates that government can take the lead.

Juliana Rotich of Ushahidi also notes that governments must invest in applications that make the data they are releasing useful. "Beyond

providing the data online, governments need to invest in applications that make the data accessible and useful to citizens," she says. Rotich cites the city of Chicago's Open Data portal as a notable example

Private Sector Development

- Once proper regulations are in place and public trust about the use of data has been gained, telecoms can compile or 'curate' mobile-generated data for use by both profit-seeking enterprises and development organisations

of a government opening data in the context of a robust multi-sector effort.

Call to Action

To realise the mutual benefits of creating an environment for sharing mobile-generated data, all ecosystem actors must commit to active and open participation. Governments can take the lead in setting policy and legal frameworks that protect individuals and require contractors to make their data public. Development organisations can continue supporting governments and demonstrating both the public good and the business value that data philanthropy can deliver. And the private sector can move faster to create mechanisms for the sharing of data that can benefit the public.

Despite the challenges and risks, the opportunities

available to better serve individuals in emerging markets should outweigh these risks. ■

For more information please visit:

World Economic Forum Mobile Financial
Services Development Report
www.weforum.org/mfs

Global Viral Forecasting
www.gvfi.org

UN Global Pulse
www.unglobalpulse.org

Ushahidi SwiftRiver Platform
ushahidi.com/products/swiftriver-platform

The World Economic Forum
91-93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland
Tel.: +41 (0)22 869 1212

© 2012 World Economic Forum All rights reserved.

No part of this publication may be reproduced or
transmitted in any form or by any means, including
photocopying and recording, or by any information
storage and retrieval system.

ⁱ www-01.ibm.com/software/data/bigdata/

ⁱⁱ www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html

ⁱⁱⁱ ushahidi.com

^{iv} www.itu.int/ITU-D/ict/facts/2011/material/ICTFactsFigures2011.pdf

^v ess.santafe.edu/publications.html

^{vi} www.unglobalpulse.org/projects/twitter-and-perceptions-crisis-related-stress

^{vii} www.nextbillion.net/blog/the-age-of-big-data

^{viii} jana.com/research/sample/,
www.nextbillion.net/blog/2011/10/05/reaching-the-next-billion-through-mobile

Haiti disaster example reference is at Karolinska Institutet
(www.ki.se)