

The Benevolent Side of **BIG DATA**

Christopher W. Surdak and Sara Agarwal

A farmer uses a simple mobile phone to obtain crop prices and subsequently finds a buyer who will pay the most: this is the most common example of how technology can spur social and economic advancement.

But today's technology allows us to take that example up a notch. With a smartphone and simple apps, buyers can track how many farmers they see in a day, and farmers can keep tabs on how much they make in a year. With a year's worth of collected information an app can let farmers know the best day and location to sell their goods. Applications link market demand with weather patterns so farmers know the best day to plant and harvest. Such analytic tools also help farmers understand precipitation and soil conditions and how these will affect their crops. They may also alert farmers to the spread of crop diseases and how to prevent them—with the click of a button. Through the use of big data, small-scale farmers can make better decisions that lead to even greater efficiency and profitability.

Big data has often been defined as the collection, analysis, and use of vast quantities of data. But existing tools already do that; what makes big data different is the ability to discover previously undetected relationships through large and sometimes unrelated sources. We can now ask new questions and get new answers. And we can do it via shared infrastructure (such as cloud computing), so almost anyone can use it. We can learn more about human behavior than ever before, and behavior is strongly linked to problem solving.

As with all technologies, big data does have a dark side. The collection of vast quantities of data can enhance people's lives, but it can also be used to control, manipulate, and exploit. Most organizations use information with good intentions. Some do not. There is growing concern over how people are being

influenced by data, and many governments are waking up to the need for regulation.

Realizing the promise of big data in developing economies requires a revolution both in the use of technology and its application. Only significant change on both fronts will allow big data to fully aid development. Governments, along with nongovernmental organizations (NGOs) and the donors and foundations that support them, have a critical role to play.

Data analytics can be used to drive growth in the developing world



Potato farmer using smartphone, Munshigonj, Bangladesh.

Big data is rapidly becoming an exercise in asking new and better questions. Each time you use your phone or post to a social media platform, you generate data that accumulate in a broader digital sea. Analysts can harvest these data, and instead of just asking *what* happened, they can now pose the

We can now ask new questions and get new answers.

more powerful question *why*. Why do people do what they do? Why did a particular input generate a particular result? The answers to these questions can reveal a great deal.

For example, in an attempt to understand why a particular region has more fatal health problems per capita, an analyst might find that fewer people in that region see a doctor at certain times of the year. A cross-check with other data reveals a correlation between doctor's visits and lower water levels at that time of year. When water levels are low, enemies from a bordering region can cross into the area of the regional hospital. Fearing for their lives, people avoid the hospital, even if they are sick. Placing a clinic in another area would resolve this issue, answering an old problem in a new way.

Practical solutions

HP Earth Insights, a program developed jointly by Conservation International and Hewlett-Packard Company, vastly improves the collection, analysis, and sharing of data on vegetation, species, carbon stocks, and climate. Scientists used to spend weeks poring over disparate data and huge numbers of seemingly disconnected photos of animals in the wild in 17 tropical forests across 15 countries. Using big data software to integrate diverse data sets and run computations and simulations, they can now identify trends in minutes, and better understand how forest loss and climate change lead to loss of species. Conservation International can then propose timely solutions and see what works best.

Or consider Kreditech, a private company that uses big data to assess the credit record of potential borrowers through publicly available information, including social networking sites. This is relevant because few people in developing economies have a traditional credit record, limiting their ability to secure a loan at all or at reasonable interest rates. But Kreditech's alternative methods can promote the granting of unsecured loans to those previously deemed unqualified. So, for example, if an applicant's friend had borrowed successfully, that would be considered a positive factor in the applicant's own creditworthiness. So far, Kreditech's repayment rate in some markets is as high as 93 percent, suggesting their criteria are not only legitimate, they are potentially better than traditional metrics. Being able to work outside traditional credit bureaus means millions of people who do not have a credit score may soon be eligible to receive credit.

Additional examples of big data influencing development include French multinational telecommunications corporation Orange and Swedish nonprofit Flowminder Foundation, both of which are using mobile phone voice and text data to

monitor population movement so they know where Ebola treatment centers would be most effective.

Big data can improve the analysis of existing development problems, but only if bright minds tackle these questions and lead the effort to address social and economic problems. Here, contextual information is the difference between discovering facts that are interesting and those that can be acted on locally. It is not enough to import data scientists from abroad. Local people familiar with local problems and culture are better placed to analyze the data. International donors and governments therefore have a key role to play in supporting the education of local data analysts. In turn, policymakers must be willing to experiment with the creative and daring new policy solutions that big data enables.

Technology booster shot

Many people wonder whether big data can fill gaps where information from traditional government data sources is incomplete or inaccurate. We can already learn much from data available in developing countries through phones, electric meters, and satellites (for example, travel patterns or economic activity). But much more could be done with better infrastructure.

Additional data can be harvested from an upgraded version of the cell phone that most people in emerging markets already have. As more and more people trade in their cell phones for smartphones, the pool of relevant data will grow. This is because smartphones encourage more interactive behavior via apps, creating a richer trove for analysts to mine. Encouraging the adoption of smartphones to address problems in specific sectors (like health and agriculture) and locations (like certain sized towns) would mean a geometric increase in both information quantity and quality.

Donors and governments have been reluctant to invest in smartphones for people who cannot afford them, as they may be perceived as a luxury item, like a television. But, given the value to society of the data generated, governments should consider subsidizing either the cost of smartphones or the associated data plans in order to collect user data. The costs may be recovered through the ethical selling of the anonymized data to those who seek them for commercial purposes. In the data-collection world, smartphones are not a luxury; they are the last mile of critical public infrastructure.

Once more people have smartphones, additional apps will be needed to produce relevant local content. Apps that help people locate clean water close to home, warn of an outbreak of cholera or Ebola, or help them to support their local school will in turn yield more information to fuel even better data analysis—and answer questions we haven't even thought to ask. Indeed, we will not know the full extent of what can be achieved until such apps are in use.

Apps created and managed by local people who have direct knowledge of the needs of their neighbors will be essential. Local developers will be key to this effort, and their initiatives should be actively encouraged. Greater use of smartphones will lead to a market for more application developers, and governments, donors, and nongovernmental organizations

can provide training, workspace, and other modes of support. The work of software developers should be seen as an extension of official statistical office work: their role in information collection is equally vital.

Some people worry that big data will be used in emerging and developing economies to shore up authoritarian governments that might use this information to maintain power. At a minimum, violations of privacy can readily occur from the misuse of collected data. But it is what we do with any technology—not the technology itself—that determines whether or not it is good or bad.

Privacy policies and other safeguards are necessary to build confidence in data collection and ensure its appropriate use. Data collected now sit with mobile operators, governments, app providers, and social media network providers. Opening up those data in an aggregated, anonymized fashion for data analysis is critical, and governments should play a role in shaping policies to enable it (even for private providers) when there is a legitimate need.

Maximizing mosquito nets

Some will argue that big data tools and the investment they require are not as important to developing economies as, say, mosquito nets. But what if big data could better inform the timing of mosquito net distribution? Indeed it already can. aWhere's Mosquito Abatement Decision Information System combines satellite data imagery, weather models and mosquito biology to identify potential onslaughts before the

mosquitoes even hatch—ultimately giving mosquito net providers and users invaluable information about where supplies can best be used.

This example showcases how better use of data can enhance existing development efforts by ensuring the optimal use of scarce resources. Big data is not just about greater awareness or more accurate forecasts; it is also about doing more with less.

Using analytics to drive growth in the developing world requires the following steps: invest in the right technology and education, support entrepreneurs and a shift in mindset, and implement privacy and open-data policies. These efforts are best implemented simultaneously because a shortfall in any area will stifle the broader effort.

Big data can and will be relevant in developing economies, but how quickly and how effectively depends on what we do now. To realize big data's full potential, we must enable both new information sources and new thinking about how to use them. Only then can we lift up more of the world's developing economies with the use of this powerful new tool. ■

Christopher W. Surdak is the author of Data Crush: How the Information Tidal Wave Is Driving New Business Opportunities and works in the field of electronic discovery and analytics, and Sara Agarwal is the Director for International Finance Organizations at Hewlett-Packard Company.



PROGRAM IN ECONOMIC POLICY MANAGEMENT (PEPM)

Confront global economic challenges with the world's leading economists, policymakers, and expert practitioners, including Jagdish Bhagwati, Guillermo Calvo, Robert Mundell, Arvind Panagariya, and many others.

A 14-month mid-career Master of Public Administration focusing on:

- rigorous graduate training in micro- and macroeconomics
- emphasis on the policy issues faced by developing economies
- option to focus on Economic Policy Management or International Energy Management
- tailored seminar series on inflation targeting, international finance, and financial crises
- three-month capstone internship at the World Bank, IMF, or other public or private sector institution

The 2015–2016 program begins in July of 2015. Applications are due by January 5, 2015.

pepm@columbia.edu | 212-854-6982; 212-854-5935 (fax) | www.sipa.columbia.edu/academics/degree_programs/pepm
To learn more about SIPA, please visit: www.sipa.columbia.edu