



ROGER LEMOYNE/REDUX/EYEVINE

Young Rohingya refugees in Bangladesh in 2018.

# Don't forget people in the use of big data for development

Information from satellites, mobile phones and other sources is not a panacea for international-development problems, warns **Joshua Blumenstock**.

Today, 95% of the global population has mobile-phone coverage, and the number of people who own a phone is rising fast (see 'Dialling up')<sup>1</sup>. Phones generate troves of personal data on billions of people, including those who live on a few dollars a day. So aid organizations, researchers and private companies are looking at ways in which this 'data revolution' could transform international development.

Some businesses are starting to make their data and tools available to those trying to solve humanitarian problems. The Earth-imaging company Planet in San Francisco, California, for example, makes its high-resolution satellite pictures freely available after natural disasters so that researchers and aid organizations can coordinate relief efforts. Meanwhile, organizations such as the World

Bank and the United Nations are recruiting teams of data scientists to apply their skills in statistics and machine learning to challenges in international development.

But in the rush to find technological solutions to complex global problems there's a danger of researchers and others being distracted by the technology and losing track of the key hardships and constraints that are unique to each local context. Designing data-enabled applications that work in the real world will require a slower approach that pays much more attention to the people behind the numbers.

## THE PROMISE

Already, data from mobile phones have transformed consumer lending in many developing countries<sup>2</sup>. Around five years

ago, researchers discovered that some people, such as those who frequently make international calls or who have more Facebook friends than others in the same area, are more likely to repay their debts<sup>3,4</sup>. Machine-learning algorithms can detect these patterns and spit out credit scores for hundreds of millions of people who own a mobile phone, but who would otherwise be shut out of formal financial services because they lack collateral or access to a bank.

Other studies have shown that, with some tweaks, the same algorithms that Google, Facebook and other companies use to match advertisements to people online can also be used to match resources to people living in poverty<sup>5,6</sup>. These algorithms identify the 'digital signatures' of poverty in personal data from mobile-phone

networks and in imagery from satellites. For instance, in most African countries, wealthier people tend to make more international calls than poorer people, who in turn are more likely than wealthier ones to live in houses with thatched roofs — as seen in satellite images. Studies from the past couple of years show that related approaches can be used to generate high-resolution maps of crop yields and childhood malnutrition<sup>7,8</sup>.

In principle, such maps could enable governments and others to distribute humanitarian aid in a much more focused and timely way than they generally do now. Analysts estimate that the wealthy benefit more than the poor in one-quarter of all interventions aimed at reducing poverty. In two projects started in Armenia in 1996, only 8% of the tens of millions of dollars earmarked for the country's neediest citizens actually reached them<sup>9</sup>.

Analysis of people's digital footprints could similarly improve public-health interventions during an epidemic, or assist national and international responses to crises. For instance, researchers have used phone data to reveal which neighbourhoods and individuals are most affected by natural disasters, where people relocate to, and how relocation affects the spread of disease<sup>10–12</sup>. Within a few years, it should be possible to track the effects of a natural disaster on individuals minute-by-minute — much as investors track the fluctuations of their stock portfolios.

### THE PITFALLS

There are at least four problems with such tools.

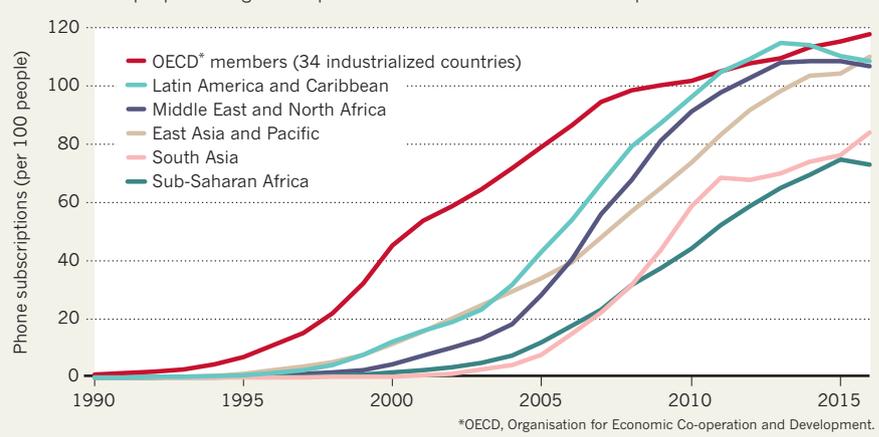
**Unanticipated effects.** Solutions enabled by big data often bolster those who are already empowered rather than vulnerable people — largely because the power to derive value from the data tends to be concentrated in the hands of a few.

Take the example of 'digital credit'. Would-be borrowers are assessed using credit scores based on their history of phone use, and loans are dispatched instantaneously by mobile phone. A booming industry has developed since the first such service, M-Shwari, was launched in Kenya in 2012. Banks, phone companies and next-generation financial-service providers collectively make hundreds of thousands of loans per day in sub-Saharan Africa alone. Today, more than 25% of the Kenyan population has taken out at least one digital loan (see [go.nature.com/2jytdp2](http://go.nature.com/2jytdp2)).

So far, to my knowledge, no published study has documented whether these loans help people, or whether — like many short-term, high-interest 'payday' loans in the United States — they lead to poverty cycles and debt traps or prevent people from later

### DIALLING UP

The number of people owning mobile phones worldwide has soared over the past decade.



obtaining loans from a bank as result of missing payments. Indeed, a substantial literature on microcredit, which predates digital credit by several decades, indicates that not everyone benefits from being able to borrow money<sup>13</sup>.

Certainly, most digital-credit customers are first-time borrowers, and surveys suggest that many don't understand the terms of the loans they are being offered. For instance, a 2015 study in Rwanda found that only 51% of borrowers were aware of the interest rate they were being charged<sup>14</sup>.

Risks of misappropriation extend beyond companies. Indeed, the potential for technology to be used in ways that don't necessarily benefit citizens could be much greater in countries where social institutions are weak and semi-authoritarian regimes more common. For example, various reports from China suggest that some people are being blocked from using trains and aeroplanes because of low 'social credit scores' — including those who have reportedly spread false information about terrorism or committed financial wrongdoings (see [go.nature.com/2wwcnwq](http://go.nature.com/2wwcnwq)).

**Lack of validation.** Conventional data-collection methods in international development, which involve surveys and face-to-face interviews, are imperfect. But they have been developed over decades, and their limitations are well documented. By contrast, the flaws in the new approaches are not well understood. There is a risk that such tools will be deployed before they have been adequately tested.

With digital data, granular maps of the distribution of wealth in a nation can be produced at a fraction of the cost of a conventional household census. But the accuracy of

such maps has been tested in only a handful of countries. And evidence suggests that patterns detected in one place do not always generalize. The tendency to make a lot of international phone calls correlates with wealth in Rwanda more strongly than it does in Afghanistan, for example<sup>15</sup>.

More worrying is the lack of evidence that such algorithms will remain accurate over time. With colleagues, I have been working on interactive tools to provide real-time visualizations of population poverty and vulnerability. By benchmarking predictions with multiple rounds of survey data (including responses to questions about income, health and employment status), we've seen that the accuracy of our maps degrades quickly, sometimes within just a few months. Why might a model trained to predict wealth from phone data collected in the winter perform poorly in the summer? Because the relationship between poverty and phone use can change. For instance, wealthier people might make more international calls than poorer people during the holiday season, but that pattern could shift during the pilgrimage months, when many more people overall are travelling.

Finally, when people become aware of the fact that their personal data are being monitored to make decisions — for instance, about who gets humanitarian aid or who is eligible for a loan — they are inevitably incentivized to game the system. GiveDirectly, a non-profit organization in Africa and the United States that enables direct cash transfers to people living in poverty around the world, initially used satellite imagery to target aid to households with thatched roofs. But people soon caught on, to the point at which some would pretend to live in a thatched structure adjacent to their main iron-roofed house to become eligible for the aid.

**Biased algorithms.** When tools are trained on biased or patchy data, those who are

*"Solutions enabled by big data often bolster those who are already empowered."*

poorly represented are often marginalized. This can be especially problematic for people in emerging economies: globally, the most disadvantaged people tend to be the least represented in new sources of digital data.

Representation can vary considerably even within such nations. Data from navigation apps such as Google Maps or Waze, for example, are increasingly being used to understand urban mobility<sup>16</sup>. But such apps typically require a smartphone, so any policy decisions made on this basis might primarily benefit the wealthier segments of society.

A mobile phone requires connectivity and electric power. Engaging with social media requires some literacy. And many digital-credit platforms require people to have a smartphone and a Facebook account. These prerequisites exclude vast segments of the population in developing countries.

**Lack of regulation.** Conventional development data are usually collected and disseminated by government agencies and aid organizations. The data that underpin artificial-intelligence applications are generally owned and controlled by private companies, who have little incentive to do anything except maximize profits.

In most wealthy countries, legislation is intended to limit governments' and companies' abuse of power. The US Supreme Court recently ruled that law-enforcement agencies cannot access phone data without a warrant. The General Data Protection Regulation in Europe is even more restrictive. In many developing nations, few such checks and balances exist, and those regulations that do exist are seldom enforced<sup>17</sup>. Currently, issues of data privacy, algorithmic transparency, fairness and accountability are off the radar of most companies operating in developing countries.

## WAYS FORWARD

Several steps can be taken to try to address these concerns.

**Validate.** New sources of data should complement, not replace, old ones. Conventional data sets are essential to calibrate and validate big-data applications. And when tools such as poverty maps are used, they need to be benchmarked against existing methods.

An example of this two-pronged approach is ongoing work to evaluate the World Food Programme's aid efforts in Haiti. The organization is aware of the potential to save costs by using phone data. So, in collaboration with researchers, it is running a comparison, collecting phone and survey data side-by-side.

**Customize.** In most cases, the core technology being used is one that's been designed for a first-world purpose — say, automatically tagging Facebook photos

with the names of friends. Finding that an algorithm can be adopted for a different use, such as identifying pockets of poverty in satellite photos, is a crucial insight — but further customization is needed before it can be useful to policymakers on the ground.

In the case of digital credit, for example, a learning algorithm might be remarkably accurate at predicting loan repayment, but the lending decision should also take local context into account. With this in mind, my colleagues and I are collaborating with Branch, a California company that provides micro-loans to millions in Africa, to research algorithms that weigh the borrower's default risk against the probable impact of a loan. The idea is to incorporate, from the start, a way to identify whether the loans are actually beneficial and to link each borrower to an 'impact score' as well as a 'credit score'. Insights emerging from the community of researchers focused on making machine-learning algorithms more fair, accountable and transparent (FAT) should help.

**Deepen collaboration.** Much of the innovation comes from the private sector — specifically from engineers in Silicon Valley in California. Many companies, including those pioneering digital credit, are motivated by the desire to do good as well as the promise of large profits.

But development challenges cannot be tackled as '20%-time' side-projects. And next-generation solutions must be designed and produced by people who understand the problems and context — not just by those who understand the algorithms.

One way to achieve this is to foster collaboration between data scientists, development experts, governments, civil society and the private sector — and especially with people and organizations in the country in question. A step in the right direction is DataKind, a global network that attempts to link data scientists to social-change organizations, many of which are focused on issues in emerging economies. As are the Data for Development challenges: in 2012 and 2014, the Paris-based phone company Orange made troves of data available to researchers all around the world, which enabled early work on poverty mapping and urban planning. Meanwhile, fellowships, competitions, internships and perhaps some kind of year abroad offered to engineers in Silicon Valley could improve data scientists' understanding of the challenges people face in different countries.

Better still are efforts to increase technical capacity locally. Google and Facebook are funding an African Master's of Machine Intelligence: a one-year intensive programme that launches in Rwanda this month (see [go.nature.com/2mcxjpc](http://go.nature.com/2mcxjpc)). Also

encouraging are the three-week intensive summer boot camp now offered at the University of Cape Town in South Africa, and Data Science Africa, a conference conceived by African researchers. Likewise, the iHub, a creative work space in Nairobi, has helped to incubate hundreds of Kenyan start-ups. But these are exceptions, not the norm. Many more such efforts are needed.

## A HUMBLER DATA SCIENCE

I am among those who are convinced that big data could transform international development. But many alleged silver bullets have missed their mark in recent decades. Think of the one laptop per child initiative. Hailed as a world-saver (see [go.nature.com/2lgoaqh](http://go.nature.com/2lgoaqh)), the technology fizzled out because developers failed to understand the social and cultural environment in which it was rolled out<sup>18</sup>.

Mike Driscoll, chief executive of the platform Metamarket, has described data science as "a blend of Red-Bull-fueled hacking and espresso-inspired statistics". In my view, the successful use of big data in development requires a version of data science that is considerably more humble than the one that has captured the popular imagination. ■

**Joshua Blumenstock** is director of the Data-Intensive Development Lab and assistant professor at the School of Information, University of California, Berkeley, USA.

e-mail: [jblumenstock@berkeley.edu](mailto:jblumenstock@berkeley.edu)

1. GSM Association. *Unlocking Rural Coverage* (GSM, 2017).
2. Francis, E., Blumenstock, J. & Robinson, J. BREAD Working Paper No. 516 (2017).
3. San Pedro, J., Proserpio, D. & Oliver, N. in *User Modeling, Adaptation and Personalization* (eds Ricci, F. et al.) 195–207 (Springer, 2015).
4. Björkregren, D. & Grissen, D. *Behavior Revealed in Mobile Phone Usage Predicts Loan Repayment* (SSRN, 2015).
5. Blumenstock, J., Cadamuro, G. & On, R. *Science* **350**, 1073–1076 (2015).
6. Jean, N. et al. *Science* **353**, 790–794 (2016).
7. Burke, M. & Lobell, D. B. *Proc. Natl Acad. Sci. USA* **114**, 2189–2194 (2017).
8. Osgood-Zimmerman, A. et al. *Nature* **555**, 41–47 (2018).
9. Coady, D., Grosh, M. & Hoddinott, J. *World Bank Res. Obser.* **19**, 61–85 (2004).
10. Lu, X., Bengtsson, L. & Holme, P. *Proc. Natl Acad. Sci. USA* **109**, 11576–11581 (2012).
11. Blumenstock, J. E., Eagle, N. & Fafchamps, M. *J. Dev. Econ.* **120**, 157–181 (2016).
12. Wesolowski, A. et al. *Proc. Natl Acad. Sci. USA* **112**, 11887–11892 (2015).
13. Banerjee, A., Karlan, D. & Zinman, J. *Am. Econ. J. Appl. Econ.* **7**, 1–21 (2015).
14. CGAP and InterMedia. *Financial Inclusion Insights: National Survey Report Rwanda* (CGAP/InterMedia, 2015).
15. Blumenstock, J. E. *AEA Pap. Proc.* **108**, 72–76 (2018).
16. Hanna, R., Kreindler, G. & Olken, B. A. *Science* **357**, 89–93 (2017).
17. Taylor, L. *Environ. Plan. D* **34**, 319–336 (2016).
18. Kraemer, K. L., Dedrick, J. & Sharma, P. *Commun. ACM* **52**, 66–73 (2009).

**J.B. declares competing non-financial interests; see [go.nature.com/2tzapmu](http://go.nature.com/2tzapmu).**