

Big Data

International development as a field of research and practice has been a laggard in using big data and powerful analytics. Much of the data are of poor quality, and there are huge gaps in the information base we rely on. This situation is changing faster than anyone predicted, and the set of tools driving this evolution represent the single-most important trend in development. The proliferation of mobile technologies, computing power, and democratization of analytics within an open-source, open-data environment will fundamentally change the way we think about and do development. I provide a synopsis of the most impactful developments in three areas: the base (data) layer, the analysis layer, and the feedback (data) layer.

The Base Layer: Open Data and Big Data

How do we know what we know in the field of international development? What is the information, or evidence base? Who generates it and how? There are at least three main collectors, sorters, and repositories of development information:

international institutions (such as the United Nations (UN), World Bank, and International Monetary Fund), national and sub-national official public-sector institutions, and the private sector. Change is afoot in each.

The open-data push by international institutions such as the World Bank and African Development Bank is making a huge amount of information available to a wide range of stakeholders. Similarly, the UN's Global Pulse initiative is creating a platform to harness new data streams and stimulate collaborative inquiry.

Groups such as Development Gateway and AidData have pushed further to show what is possible by this opening. A good example is a tool like Development Loop,¹ which plots all World Bank and African Development Bank projects at precise geographic locations across Africa and overlays feedback sourced from the intended beneficiaries. This full circle or loop is a reminder of the

¹ "Development Loop App," AidData, <http://www.aiddata.org/content/index/Maps/development-loop-app>, accessed December 21, 2011.



An Iraqi woman shows an SMS text message she received on her mobile phone from the Iraqi Ministry of Health asking people to take precautions by avoiding crowded places and using handkerchiefs to avoid catching the flu. | AFP Photo: Ali Al-Saadi

importance of data transparency and universal standards. Opening up aid data in a standardized format will make geocoding a potent tool for real transparency and accountability.

To appreciate what a game-changer open data can be, consider the situation in Uganda. Research conducted there in 2006–2007 by the United Kingdom-based initiative Publish What You Fund revealed that the government was unaware of the amount donors were planning to spend for development that year. The planned expenditure was more than double what the government was aware of. Indeed, financial resources flowing into the country were far higher than had been estimated.²

² “Aid Budgets in Uganda,” Publish What You Fund, www.publish-whatyoufund.org/resources/uganda/, accessed December 21, 2011.

The World Bank’s Chief Economist for Africa calls the next example a “statistical tragedy.” Most of Africa’s population lives in countries that still use an outdated (1960s) method of national income accounting to generate fundamental data, such as gross domestic product (GDP). Ghana only shifted to the 1993 UN system of national accounts last year. When they did so, they found their GDP was 62% higher than previously thought, catapulting the country to “middle income” status.³

If even the most basic information is riddled with problems then what do we really know? Data reliability is one issue, but time lag is another. Too

³ *Blogs.worldbank.org*, “Africa’s Statistical Tragedy,” blog entry by Shantayanan Devarajan, October 6, 2011, blogs.worldbank.org/africacan/africa-s-statistical-tragedy, accessed October 6, 2011.

often, the data used in international development decisions are stale. The information base we rely on needs to be bolstered by building bridges with new sources and data streams.

Opening up proprietary private-sector data for use in international development will be a game-changer in the coming years. To date, public institutions have been leading on open data (for instance, under the purview of the Open Government Partnership). But it is the private sector, the main repository of “big data,” that is the holy grail. If you total all the data collected by the U.S. Library of Congress (one of the largest public-sector repositories), it would be about 235 terabytes as of April 2011. Walmart processes and stores about 2,500 terabytes per hour!⁴

The Analytics Layer: Virtualization, Visualization Driving Democratization

High-power analytics revolutionized the commercial sector and can now do the same in the social sector. At its core, analytics is about understanding relationships and patterns. Analytics helped retailers profit from unlikely trends, and it can do the same for complex social systems. Bringing this capacity to bear on development challenges, such as food security and urbanization, is just the beginning.

The explosion of mobile sensors—especially in the developing world—is facilitating a transformation. Mobile-phone subscriptions have grown from less than 750 million (less than a third in developing countries) at the start of the 2000s to more than 5 billion (4 times as many in developing countries as in the developed world today). About one-fifth of all subscribers live on less than \$5 a

⁴ Abhishek Metha, “Big Data: Powering the Next Industrial Revolution,” Tableau White Paper, www.tableausoftware.com/learn/whitepapers/big-data-revolution, accessed March 29, 2012.

day.⁵ The developing world is the leading driver of mobile big data. Voice, text, transactional, locational, and positional information can be overlaid with the base data layer described earlier (income, health, education, and other indicators generated by official sources) to produce new insights into real behavior and complex incentive structures.

Take the example of the Engineering Social Systems lab.⁶ Coupling terabytes of mobile-phone data with Kenyan census information, the lab is modeling the growth of slums to inform urban planners about where to locate services such as water pumps and public toilets. In Uganda, the same group is developing causal structures of food security, and in Rwanda, they used big mobile-phone data and a random survey to model how different people react to economic shocks. This constitutes a fundamental shift from theoretical models to models informed and built on real networks.

Development analysis has long been limited to correlations and inferences based on correlations. For the first time, big data coupled with high-power analytics are opening up the possibility of, if not entirely causal dynamics, then at least more robust inferences. Our traditional methods of inquiry have conditioned us to think in terms of generalizing on the basis of random sampling. But for the first time, the proliferation of mobile sensors is making possible highly targeted yet nonintrusive inquiry.

The rapid emergence of new data streams has kept pace with the development of analytical capacity to draw useful inference out of them. Twitter, for instance, generates information about

⁵ Nathan Eagle, “Mobile Crowdsourcing in the Developing World,” presentation for MIT University’s Entrepreneurial Programming and Research on Mobiles, 2008, assets.en.oreilly.com/1/event/20/tختهagle_%20Crowd-Sourcing%20on%20Mobile%20Phones%20in%20the%20Developing%20World%20Presentation.pdf, accessed March 29, 2012.

⁶ “Big Data for Social Good,” Engineering Social Systems collaboration, ess.santafe.edu/bigdata.html, accessed December 21, 2011.

the size of the entire U.S. Library of Congress in two weeks and, together with Facebook, has already shown its efficacy during the Arab uprisings. At the heart of this evolution, open-source software systems and tools allow the simultaneous collection, categorization, and analysis of various data types—from Twitter hashtags, to videos, to positional data and machine IDs. Swift River, developed by Ushahidi, is an example of a free open-source platform that enables rapid simultaneous filtering and verification of real-time data. It also visualizes the information in dashboards that the average user can understand.

This is particularly powerful for monitoring immediate post-crisis developments when the information flow suddenly increases, but it is also only useful if immediately analyzed. Similar applications were successfully implemented and yielded important insights on population movements both in the aftermath of the earthquake in Haiti as well as flooding in Pakistan.

Virtualization (of platforms) and visualization (of large complex information to make it engaging for the average user) inspire the community- or crowd-driven problem solving that advances democratization of analytics. For example, Data Without Borders, a pro-bono data scientist exchange, organizes “data dives” to help leverage the potential of information that NGOs, civil-society organizations, and others possess but do not have the time, capacity, or inclination to process.

We at The North-South Institute are playing our own small part in comprehensively visualizing Canada’s engagement with developing countries on aid, trade, investment, migration, and a range of other flows through the recently launched Canadian International Development Platform.⁷

⁷ *Canadian International Development Platform*, hosted by The North-South Institute, www.cidpnsi.ca.

The Feedback Layer: Deep Context, Complex Microsystems, Real-time Loops

The efficacy of the feedback layer is also new. Targeted crowdsourcing has already come a long way. The Ushahidi experience in Kenya, for instance, also worked for monitoring elections in Afghanistan. Mobile-phone SMS platforms have been adapted to make participatory budgeting more inclusive in hard-to-reach areas, such as conflict-affected South-Kivu in the Democratic Republic of the Congo—and results have been encouraging. The experience of the Development

A “statistical tragedy”: Most of Africa’s countries still use a 1960s method of accounting to generate fundamental data, such as GDP.

Loop initiative has already shown that, with creative use of available technologies and committed partners, it is possible to obtain direct feedback from intended recipients of interventions.

To understand how powerful the feedback layer can be, consider the experience of the Mobile Accord. At the initiative of the World Bank’s World Development Report 2011, the Accord ran Geo Poll, an SMS-based targeted polling in the Democratic Republic of the Congo. The poll asked 10 questions that included sensitive topics, such as sexual violence against women. The survey produced 1.2 million text responses, and the outputs were turned into the video “DRC Speaks,”

which captured people's responses to questions about their experiences in their own words. This ended up being one of the largest surveys ever conducted in the country.⁸

Some of the most valuable data in development come from surveys, including household, labor market, living standard, and other social surveys. But there are two key problems with such surveys: time (they take time to implement and can only be done infrequently) and high costs. Mobile technology is helping get around these issues. The World Bank is piloting an interesting initiative in Latin America called "Listening to LAC" (L2L)⁹ where several types of mobile technologies are being deployed to conduct real-time (higher frequency) self-administered surveys, to generate panel data on key questions pertaining to vulnerability and coping strategies. While still in a pilot phase, this is the first time such information is being collected near real-time and with lower costs than large national surveys.

There is a pattern here. In the base layer, more and more data are opening every day. In the analytics layer, experimental ideas are leaving the lab for real-world application. Virtualization and visualization are helping foster new communities geared toward collaborative problem solving. Similarly, in the feedback layer, tools are also democratizing. Ushahidi created an easy-to-use version of their implementation, called CrowdMap. Anyone who knows how to set up an email account can use the tool to set up their own incident mapping of whatever trend, alert, or issue on which they are interested in getting feedback from the crowd.

At sunset, a young girl tests out a new seesaw on a playground built by the Elizabeth Glaser Pediatric AIDS Foundation at the Mkhulamini Clinic in Swaziland. This year, the Foundation will launch a USAID-funded, five-year program to expand services preventing mother-to-child transmission of HIV. | Photo: Jon Hrusa, Elizabeth Glaser/Pediatric AIDS Foundation

Looking Ahead

How we think about data and analysis in the field of international development is changing rapidly, and faster than many organizations that "do development" are prepared for.

The open-data movement has widened access to a broad range of basic contextual information. A similar push is needed to open private-sector big data in the service of social good. Powerful analytical tools and collaborative platforms are dramatically changing what is possible for even the most intractable challenges like understanding socioeconomic risks and responses, dealing with urban planning, and better preparing for emergencies. For the first time, we have a feedback layer, which has made possible deep and near real-time awareness of what is working or not working, where, and why. Together, big data, democratized analytics, and the ability to tap deep contexts will change the way we think and do development in the coming years.

Aniket Bhushan is a Senior Researcher at The North-South Institute and leads the Canadian International Development Platform. The views expressed in this essay are his own, and do not necessarily represent the views of the United States Agency for International Development or the United States Government.

⁸ The World Bank, "DRC Speaks," World Development Report 2011 multimedia library, wdr2011.worldbank.org/media-library, accessed March 29, 2012.

⁹ "Getting the Numbers Right: Making Statistical Systems a Real Plus for Results," The World Bank: IBRD Results, March 2010, siteresources.worldbank.org/NEWS/Resources/Gettingthenumbersright4-19-10.pdf, accessed March 29, 2012.