

## ENABLING DIGITAL DEVELOPMENT

# The data revolution

A report commissioned by the United Nations, *A World That Counts*, proclaims, “Data are the lifeblood of decision-making and the raw material for accountability. Without high-quality data providing the right information on the right things at the right time, designing, monitoring and evaluating effective policies becomes almost impossible.”<sup>1</sup>

Recognizing the potential for harnessing the ongoing explosion of data, but mindful of gaps between the developing and developed world, the report calls for a data revolution that would aid in the achievement of the Sustainable Development Goals.

The world is witnessing an unprecedented explosion of data. Digital data overtook analog around 1998, and in 2013 amounted to 46 billion trillion bytes. That’s equivalent to about 400 trillion printed copies of this Report, which when stacked would reach from Earth to well beyond Pluto. Figure S5.1 shows the ongoing upsurge of data, as measured by total storage capacity. Figure S5.2 shows the evolution of telecommunications capacity. Although the absolute gap between higher- and lower-income countries is increasing, growth in the latter has been faster since 2008.

In harnessing this data explosion for development, attention focuses on two overlapping innovations: “big data” and open data. Big data are voluminous or fast. They come, for instance, from satellite and ground sensors and as by-products (“data exhaust”) from electronic transactions and from mobile phone calls. The promise of big data is to provide information of unprecedented scope, detail, or rapidity. For instance, Global Forest Watch crunches massive amounts of open satellite data in order to generate near real-time, global maps of tropical deforestation.<sup>2</sup>

Open data are those that are freely and easily accessible, machine-readable, and explicitly unrestricted in use. Open data aren’t necessarily big, and big data aren’t necessarily open. Governments are, or could be, important sources of data on population, public budgets, education and health facility usage and status, weather, and trade. When opened, these data can be combined and recombined in ways that directly benefit the public (for instance, by increasing the transparency and accountability of government) and provide the basis for commercial, value-added services (such as apps for navigating public transit). Box S5.1 presents some examples of both big and open data.

Exuberant estimates of the current and potential economic value of big data and open data range from the hundreds of billions to the trillions of dollars per year. The clearest benefits so far have been for map and weather data. Openly available global positioning system (GPS) data supported markets for geospatial data and services worth US\$56 billion in 2013 in the United States alone.<sup>3</sup> At least four companies valued at over US\$1 billion—Zillow, Zoopla, Waze, and the Climate Corporation—process and resell open data about real estate, traffic conditions, and weather.

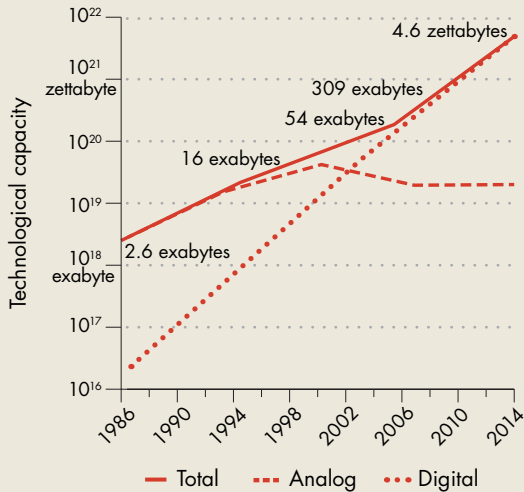
Yet sustained, impactful, scaled-up examples of big and open data in the developing world are still relatively rare. Open data has far to go. Figure S5.3 shows that readiness, implementation, and impact of open data are all highly correlated with GDP per capita, but that there are shortcomings in high-income countries, as well.

Data impact requires willing suppliers and eager demanders. On the supply side, private holders of data may be reluctant to share it for fear of jeopardizing customers’ privacy or corporate competitiveness. Yet some of these data, if pooled and shared, could

WDR 2016 team, incorporating contributions by Amparo Ballivian and Thomas Roca (2015).

**Figure S5.1 World's capacity to store information**

in optimally compressed bytes

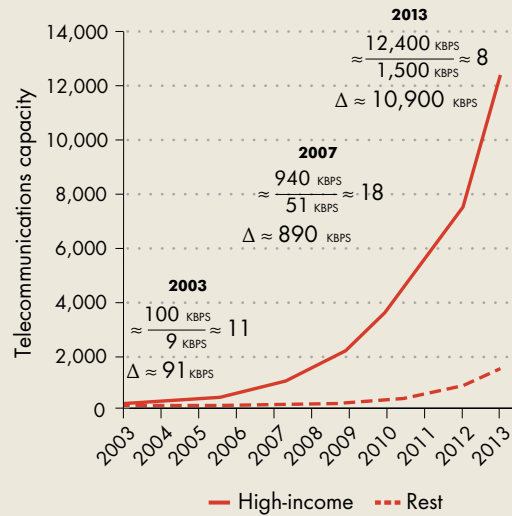


Source: Hilbert 2015. Data at [http://bit.do/WDR2016-FigS5\\_1](http://bit.do/WDR2016-FigS5_1).

generate public benefits. There are emerging models that seek to create a symbiotic relationship among public, private, and academic actors that incorporates benefits for all.

**Figure S5.2 Growth in telecommunications capacity**

in optimally compressed kbps



Source: Hilbert 2015. Data at [http://bit.do/WDR2016-FigS5\\_2](http://bit.do/WDR2016-FigS5_2).

Note: kbps = kilobits per second.

For instance, taxi companies may justifiably be unwilling to share detailed data on vehicle movements. Data on average vehicle speed by road segment is valuable for the companies—it helps them

**Box S5.1 “Big data” and open data in action**

**Informing relief efforts in the wake of the Nepal earthquake.** A critical need in disaster relief is to track displaced populations for efficient logistics planning. Cellphone location data can provide comprehensive, real-time information on population, but cellphone operators are often reluctant to share this data for technical, confidentiality, or competitive reasons. Flowminder, a Swedish nongovernmental organization (NGO), has worked out procedures for accessing this data and has used it to estimate population movements following the 2015 earthquake to aid in relief efforts. <http://www.worldpop.org.uk/nepal/>.

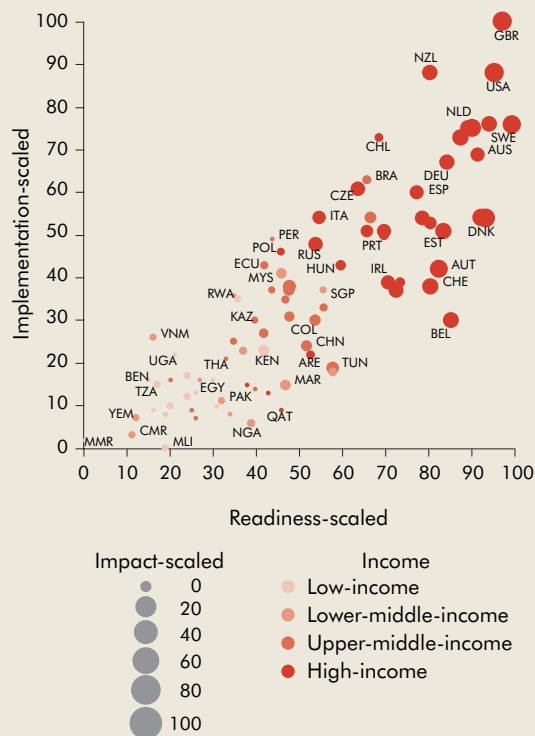
**Real-time independent measures of inflation.** PriceStats computes daily inflation data for 22 economies by scraping price data from the web. These inflation statistics are more timely than official numbers, and provide an independent cross-check. <http://www.pricestats.com>.

**Accountability for subsidies in Mexico.** Fundar Center for Analysis and Research, a Mexican NGO, persuaded the Mexican Ministry of Agriculture to open its data on the large PROCAMPO subsidy program. The data showed that 57 percent of the benefits were going to the wealthiest 10 percent of recipients. A website now tracks these and other financial flows, and allows data to be visualized. <http://subsidiocalcampo.org.mx/>.

**Promoting efficient procurement of pharmaceuticals in Southern Africa.** The Southern African Regional Programme on Access to Medicines and Diagnostics (SARPAM) InfoHub assembles information about pharmaceutical procurement volume and prices in the Southern African Development Community. It publishes price data on the web, and estimates the potential savings, for each member country, if medicines were purchased at the lowest available price. <http://med-db.medicines.sadc.int/>.

Source: WDR 2016 team.

**Figure S5.3** Readiness, implementation, and impact of open data



Source: WDR team based on 2014 data from *World Wide Web Foundation 2015*. <http://barometer.opendataresearch.org/report/about/data.html>. Data at [http://bit.do/WDR2016-FigS5\\_3](http://bit.do/WDR2016-FigS5_3).

with dispatching—and for city managers it provides a useful gauge of congestion. The World Bank is exploring ways to provide companies with software that distills their data down to average vehicle speed. This summarized data could then be shared between companies and with the public, yielding more accurate and comprehensive estimates of travel speeds.

In Australia, a tech startup called The Yield is working with public sector regulators to test ways to help farmers share data for the common good. The \$A 100 million Australian oyster industry loses nearly \$A 34 million a year due to harvesting closures by regulators to protect human health when runoff is contaminated. The idea is to equip the oyster farms with real-time sensors on water quality. The data will help individual farmers with farm management, while allowing regulators to optimize decisions, avoiding costly, unnecessary closures and still maintaining safety.

Public agencies are also reluctant to share data, even when it has large public benefits. For example, of the 86 countries surveyed by the Open Data Barometer, one-third of the high-income countries and 85

percent of the remainder had made little or no progress in opening map data. One reason is that cash-strapped agencies support themselves by selling data. This shuts out many potential users, even though the additional cost of providing the data to them would be negligible. A dramatic example of reversing this trade-off is the decision of the U.S. government to reduce the cost of a digital Landsat satellite image from US\$600 to zero in 2008. The annual number of downloads increased from fewer than 25,000 to 3 million. The direct benefits are estimated at US\$1.8 billion a year for U.S. users and US\$400 million for international users;<sup>4</sup> there are substantial indirect benefits as well. The loss of government revenue from selling the 25,000 images was miniscule by comparison. So if the fixed costs of data maintenance can be financed, the benefits from free distribution can be large.

There are other reasons for public agencies to be reluctant to share. Agencies may lack technical skills to share data. They may be unwilling to expose the data to scrutiny because of quality concerns. They may value the political leverage afforded by monopolizing data. Nonetheless, both the G-8 countries and the African Union's High-Level Conference on the Data Revolution have endorsed the concept of making official data open by default.

Effective demand for data may also be weak, even if there is underlying interest. Of course, accessibility to the internet is a prerequisite. Outdated or poor-quality data will discourage demand. Because data literacy is critical, intermediaries can play a key role in interpreting, analyzing, and disseminating open data. But capacity or incentive to do this is often lacking. There has not been as much uptake by journalists as some have hoped. And while a benefit of open data is enhanced sharing among government agencies, relatively few governments are making optimal use of existing statistics for planning purposes, according to PARIS21. However, Bangladesh, Cambodia, and Niger scored high on statistical capacity (according to PARIS21's IDR Metadatabase), showing that it is possible for low-income countries to excel at generating and using statistics.<sup>5</sup>

The pioneering Kenyan Open Data Initiative shows both the promise and challenges of open data. A champion, Permanent Secretary Bitange Ndemo, persuaded ministries to open their data, and the program was inaugurated in 2011. The World Bank, Google, and Ushahidi provided technical and financial support. Drawing on Nairobi's vibrant tech scene, the ancillary Code4Kenya initiative spurred the development of apps that presented health, safety, and other information in engaging and useful ways.

A journalist, Irene Choge, attended a Data Journalism Boot Camp. She then used the Open Data Platform to trace a chain of links: funding for school toilets had gone missing, so the student-to-toilet ratio for many schools was intolerably high, leading to open defecation; consequently students were suffering from illnesses associated with poor sanitation, and female students, especially, were discouraged from attending school; the end result was poor performance on tests. The findings prompted government action.<sup>6</sup>

Despite this promising beginning, the Open Data Initiative lost steam. Data were not updated, the apps struggled to find users, and a legal framework for open data was not in place by mid-2015. Kenya's score on the Open Data Barometer dropped from 43 to 26 between 2013 and 2014.

This example underscores the global need for sustained commitment to openness and for investment in high-quality data if the dream of a data revolution is to be realized. Countries can seek ways to discourage data hoarding, by adequately funding data holders and showing them that they can attain a higher profile by opening their data. Donors and the international community can support these actions through funding, capacity building, and ensuring that their own data are open.

## Notes

1. Independent Expert Advisory Group Secretariat 2014.
2. <http://www.globalforestwatch.org/>.
3. <http://gpsworld.com/the-economic-benefits-of-gps/>.
4. Miller and others 2013.
5. <http://datarevolution.paris21.org/metabase>.
6. <https://www.youtube.com/watch?v=A58R2yNQtio>.

## References

- Ballivian, A., and Thomas Roca. 2015. "New Data for Development." Background paper for the *World Development Report 2016*. World Bank, Washington, DC.
- Hilbert, Martin. 2015. "Quantifying the Data Deluge and the Data Drought." Background note for the *World Development Report 2016*. World Bank, Washington, DC.
- Independent Expert Advisory Group Secretariat. 2014. *A World That Counts: Mobilizing the Data Revolution for Sustainable Development*. New York: United Nations.
- Miller, Holly M., Leslie Richardson, Stephen R. Koontz, John Loomis, and Lynne Koontz. 2013. "Users, Uses, and Value of Landsat Satellite Imagery—Results from the 2012 Survey of Users." Open-File Report 2013–1269. U.S. Geological Survey. <http://pubs.usgs.gov/of/2013/1269/>.