April 2015

Development Progress

Research Report 03

The data revolution

Finding the missing millions

Elizabeth Stuart, Emma Samman, William Avis and Tom Berliner

Overseas Development Institute
203 Blackfriars Road
London SE1 8NJ

The Institute is limited by guarantee
Registered in England and Wales
Registration no. 661818
Charity no. 228248

Contact us
developmentprogress.org
developmentprogress@odi.org.uk
T: +44 (0)20 7922 0300

Sign up for our e-newsletter
developmentprogress.org/sign-our-newsletter

Follow us on Twitter
twitter.com/dev_progress

Disclaimer
The views presented in this paper are those of the author(s) and do not necessarily represent the views of ODI.

Cover image: Girl on mobile phone at community meeting in Aurangabad, India.
© Simone D. McCourtie / World Bank.

Corrections *(29/05/2015)*

Infographic one
Amended to include the estimate and upper estimate of maternal deaths, and to remove lower estimate, with the text amended accordingly.

Infographic two
Updated to focus on a lower estimate of the extent to which poverty may be underestimated based on those missing from household surveys.

Infographic three
Updated to reflect a longer time window and availability of two estimates necessary to record trends.

Page 12, paragraph 4
This has been clarified to reflect the fact that considerably more data are now available for the 2006-2013 period than was the case at the time the Chandy (2013) paper was written.

Page 13, paragraph 1
Paragraph and footnote updated to reflect the fact that more recent poverty data are now available for Botswana.

Page 13, paragraph 3
Amended to read: 'poverty figures could be understated by at least one-quarter'. This is to reflect that the numbers are in-house computations of Carr-Hill (2013) data.

Contents

List of tables, figures and boxes

Tables

Figures

Boxes

Acknowledgements

This case study was written by Elizabeth Stuart, Emma Samman, William Avis and Tom Berliner from ODI.

The authors would like to thank Sabina Alkire (OPHI), Alex Cobham (CGD), Graham Eele (independent consultant), Ricardo Fuentes-Nieva (Oxfam), Grace Kite (Gracious Economics), Claire Melamed (ODI), Sheila Page (ODI), Alex 'Sandy' Pentland (MIT), Louise Shaxson (ODI), Kevin Watkins (ODI) and Leni Wild (ODI) for their detailed comments on and peer reviews of earlier drafts of this paper. Mistakes, however, remain the authors' own. The authors also extend their gratitude to Juan Carlos Barahona, Owen Barder, Sylvia Beales, Thomas Bjelkeman-Pettersson, Grant Cameron, Gero Carletto, John Clippinger, Andrew Dilnot, Amanda Glassman, Mark Graham, Trevor Fletcher, Vanessa Frias-Martinez, Natasha Frosina, Rebecca Furst-Nichols, Paula Hidalgo-Sanchis, William Hoffman, Daniel Jiminez, Amina Khan, Robert Kirkpatrick, Sunita Kishor, Petra Krylova, Pali Lehohla, Amanda Lenhardt, Emmanuel Letouzé, Peter van den Linde, Denise Lievesley, Cosmas Ochieng, Miriam Lips, James Mansell, Matthew Martin, Robert McCaa, Raymond Mugwanya, Amanda Noonan, John Quackenbush, Guilio Quaggiotto, Rakesh Rajani, Celia Reyes, Evelyn Ruppert, Anoush Tatevossian, Chris Underwood, Simon Wanda, Firman Witoelar and attendees of the Data-Pop Alliance, ODI and Royal Statistical Society workshop on Big Data, held at the Royal Statistical Society in London, January 2015, for their input. Nina Behrman and Aaron Griffiths provided editorial support, and Chris Little and Nile Davies managed layout and production.

Responsibility for the views expressed and for any errors of fact or judgement remains our own.

The case study was funded by the Bill & Melinda Gates Foundation as part of 'Development Progress', a four-year research project that aims to better understand, measure and communicate what has worked in development and why. Its findings and conclusions are those of the authors and do not necessarily reflect the positions or policies of the Bill & Melinda Gates Foundation.

Executive summary

For governments wanting to end poverty, steward sustainable environments and foster healthy, thriving populations with the opportunity to earn a decent living, many of the necessary pieces are now in place. They start from a good base. Millions of families have escaped poverty and many million more children are in schools than was the case 15 years ago.[1] Much more is known about successful developmental pathways.[2] And many of the world's poorest countries are experiencing strong economic growth.

But, finance aside, there is still one key element the absence of which is impeding progress: data. Governments do not adequately know their own people. This is particularly true for the poorest and most marginalised, the very people that leaders will need to focus on if they are to achieve zero extreme poverty and zero emissions in the next 15 years (Granoff et al., 2014). Nor will the international community be able to support the most vulnerable and marginalised people without an overhaul of the current ways of gathering statistics. As many as 350 million people worldwide are not covered by household surveys. There could be as many as a quarter more people living on less than $1.25 a day than current estimates suggest, because they have been missed out of surveys (Carr-Hill, 2013).

Data are not just about measuring changes, they also facilitate and catalyse that change. Of course, good quality numbers will not change people's lives in themselves. But to target the poorest systematically, to lift and keep them out of poverty, even the most willing governments cannot efficiently deliver services if they do not know who those people are, where they live and what they need. Nor do they know where their resources will have the greatest impact.

To put it starkly, it is highly problematic that we set global targets to reduce poverty in all its forms without the ability to know whether or not those targets have been met.

This report sets out the evidence that, even when people are counted, the counting is frequently not good enough. What is assumed to be an empirical fact – a statistic – is too often the result not of direct observation but of inference, assumptions or extrapolation, or political negotiation. In sub-Saharan Africa, some 133,000 women may have died from childbirth-related causes in 2013, or twice as many. We cannot be sure.

We argue that good quality, relevant, accessible and timely data will allow willing governments to extend services into communities which until now have been blank spaces in planning processes, and to implement policies more efficiently, meaning that a data revolution could, in the medium term, pay for itself.

The solutions are threefold: increasing investments in the capacity of national statistical offices (NSOs), thereby potentially improving the scope and frequency of household surveys; using alternative sources of data to fill gaps and building strong administrative systems; and making better use of the data we already have. For example, the World Bank has a hidden database of consumption data that should be made open-access.

A data revolution for citizens

Perhaps even more transformative, the data revolution could play a role in changing the power dynamic between citizens, governments and the private sector,[3] building on open data and freedom of information movements around the world (see Box 13, page 36). It has the potential to enable people to produce, access and understand information about their lives and to use this information to make changes. In Colombia, big-data analytics are saving small-scale farmers $3.6 million per harvest, and increasing their independence (see Box 14, page 37).

Already, extraordinary volumes of data[4] are being produced – through social media, mobile mapping, geo-sensing and spending records – very cheaply, and often automatically. These data are very high resolution, and can give precise information at an individual level (Bellaggio, 2014). When aggregated, they can tell us things in entirely new ways, for example:

1 However, the latest UNESCO report shows that improvements in the number of out-of-school children have slowed – partly because of a failure to enrol marginalised children (UIS and UNICEF, 2015).

2 See the ODI project, Development Progress (www.developmentprogress.org).

3 Note that this paper does not focus on data and the private sector.

4 IBM estimates that we now create more than 2.5 quintillion bytes of data each day (www-01.ibm.com/software/data/bigdata/what-is-big-data.html).

- engineers can monitor water meters remotely in Liberia
- local people can predict traffic jams in Rio de Janeiro (Matheus and Manuella, 2014) and in Tallinn (Järv, 2012)
- local government can track real-time energy use in Christchurch, New Zealand (NZDFF, 2014)

People in Jakarta tweet more than in any other city. Soon, in some parts of the world, it will be possible to carry out a proxy census each week from a computer, or to track inflation on an hourly basis.

Granular data[5] can also provide information that members of the public can use to hold governments to account as in Costa Rica where all government procurement is published online, so that every bid is open to scrutiny. Of course, data alone are not sufficient to change governance overnight but, used as part of an iterative feedback process, they can help bring about necessary reforms.

A data revolution defined

The data revolution is an explosion in the volume of data, the speed with which data are produced, the number of producers of data, the dissemination of data, and the range of things on which there are data, coming from new technologies such as mobile phones and the internet of things, and from other sources, such as qualitative data, citizen-generated data and perceptions data (IEAGDRSD, 2014).

The current political context for the data revolution discussion is the Post-2015 Framework negotiation process to agree a new set of goals – the Sustainable Development Goals (SDGs) that countries around the world will commit themselves to reach by 2030. As part of this negotiation, there has been a call for a substantial investment in data because the capacity to measure and track progress is not yet there (IEAGDRSD, 2014).

However, the debate as to how best to improve data is wider than this. The value of data for the SDGs is to help governments actually achieve the goals, not just monitor that achievement. Whatever the outcome of the negotiations, a new world of data is a reality coming down the track that can be captured for the benefits of millions – or just the elite few. **There is now huge potential to fill these data gaps because new technology, increased demand for data from civil society and other actors, and the political opportunity of the Sustainable Development Goals mean that some problems may finally be fixed.**

Not all of this is about new technologies. Sometimes, the very lowest-tech solutions will create the data that respond to the needs of the poorest (Bellagio, 2014). In the remote area of Bawomataluo in South Nias district of Indonesia, simple stickers on doors showing the last time a child was weighed have been the nudge needed for people to access post-natal care (World Vision, 2014).[6] And if the data revolution is to deliver on its name, it will need to do more than just create more data, although in some cases that will be necessary. What will be far more revolutionary will be to create the right sort of data, and thereby information that will allow the right questions to be asked and answered, and to make them accessible, usable and verifiable.

Nor should it be about reinventing the wheel. Finance, health, education and environment ministries have years' worth of data printed on paper spilling from filing cabinets in yellowing sheets. Entering these into a searchable database may be far more useful than creating a new algorithm. Making better use of existing data has as much of a place in this revolution as does big data.[7] The changes outlined above are far from being an automatic outcome of the generation of new data – it would be highly naive to expect otherwise. A genuine revolution will come about only if the right kinds of data are produced, if people can use them, and if the political will is there to act upon what they tell us.

Data gaps

To put it starkly, it is highly problematic that we set global targets to reduce poverty in all its forms without the ability to know whether or not those targets have been met. This is the case at both global and national levels. Arguably, it may not be of paramount importance where Afghanistan or Burundi or El Salvador lie on an international index of access to clean water. But it certainly matters at the national level: countries set detailed national targets as part of their planning processes. While it is laudable that countries have these aims, if they have no adequate way of measuring their achievement, this risks becoming an exercise in futility.

Getting it right

The data revolution will need to be both top down, with new checks, balances and legal frameworks (and the institutional capacity to realise them), and bottom up, as citizens create, access and analyse data in innovative ways, and use those data to hold governments, the private sector and donors to account.

Getting it right will entail the creation of innovative partnerships, but also some relatively quotidian shifts to the way we gather information now. It will need to be accompanied by a high level of political will to act upon the data, and it will require application of knowledge and

5 Granular data refers to information broken down into small separate items.

6 For discussion of behavioural economics or so-called nudge theories see World Bank (2015b).

7 For discussions of big data definitions see Letouzé and Jutting (2014) and WEF (2015).

understanding of local development challenges – otherwise the data will exist in a vacuum.

It will mean that increasingly, official data collected by governments will be complemented by and combined with traditional and big data from the private sector, NGOs and individuals. A project in Uganda is using satellite data to distinguish between different types of roofs as a proxy poverty measure for example (see Box 16, page 41).[8] Government statisticians will need to act as brokers, providing quality control on data created elsewhere.

At the moment official statisticians are concerned that they do not have a methodology for integrating, say, data gathered from Twitter, with survey data. However, while rigour is important, traditional forms of data can be complemented – rather than replaced – by the speed, scale and scope of new data sources. NSOs should not be bypassed by the revolution, but instead governments need to ensure they are better funded and have more trained staff. And traditional data collection techniques – household surveys, censuses and government registers – also need to be improved.

Some caveats

While we enthusiastically embrace the data revolution, discussions around it need to be grounded in reality. It is understandable that progressive academics in New York, Silicon Valley and Nairobi tech hubs are excited by the potential of big data to change the world. But this needs to be married with the reality: that many statistics offices in developing countries have irregular electricity supply; that too often only a handful of people in the country will have the relevant training to do this work; and that the few qualified people may be pulled away from their vital work to provide answers to donors' research questions irrelevant to that country's own data needs.

Apparent quick fixes – e.g. flying in data experts from Silicon Valley to set up systems and analyse data on behalf of overstretched governments – miss the point. Either these actions prevent national statistical agencies from building their own capabilities, or, absent local knowledge or context, it leaves the new data wide open to serious misinterpretation. Yet paradoxically, the low starting point in most developing countries is the very reason for the optimism of a data revolution – done right, it can mean countries may not need to spend decades building up expensive statistical architecture.[9]

Caution is however required when considering what a data revolution can do. The best data in the world – timely, relevant and easy to understand – will not by themselves lead to change. Governments and citizens need to act upon the knowledge imparted by the data to make a difference to the lives of people – an accountability revolution. Of course, while improving accountability will in the best cases make the job of government easier by enabling more efficient targeting of service provision, for example, it would be naive not to recognise that governments and citizens too often do not have the same interests. There will be governments that actively seek to block the opening up of data, particularly data on people or elements of society they have chosen to keep in the margins.

Fears

This paper speaks to some quite justifiable fears and concerns around a data revolution. Some NSOs fear a data revolution for three reasons: first, that they, already stretched in many cases, may be overwhelmed by a new set of demands for vastly improved statistical services. Second, paradoxically, there are fears that big data may make them irrelevant – that mobile phone data analysis and automation will make censuses and household surveys obsolete. Third, there are fears that new data forms, such as big data collected from social media, will not produce methodologically sound results.

Notably, these fears are not unique to developing-country NSOs – they are shared in large part by the statistical community in other countries. Trevor Fletcher of PARIS21 says: 'I've heard people from statistics offices in rich and developing countries express concern that they will be made irrelevant and put out of business by big data, or simply won't be able to cope'. The head of the French NSO, INSEE has said that big data seems largely irrelevant to its work (Letouzé and Jutting, 2014).

There is a broader fear among some developing country governments that the data revolution is yet another donor-driven offensive, powered either by the hope that it will provide a technical fix for protracted economic, social and environmental problems, or that it is a mission to push yet more performance frameworks in disguise.

Civil society groups are concerned that the data revolution may increase inequalities; that the so-called digital divide will widen as a result. Citizens' groups and the wider public are also concerned about government surveillance, privacy and data ownership. While the above caveats and concerns are all important, and responses need to be found, we also need to avoid being so over-cautious that we miss this significant opportunity.

8 http://ssg.mit.edu/~krv/pubs/AbelsonVS_kdd2014.pdf.

9 For a discussion of the opportunity to 'leapfrog' statistical systems in developing countries, see Giugale (2012).

The data gap: how many women died in childbirth in 2013?
Even our best estimate could be nearly 40% too low.

Estimate
179,000

High estimate
256,000

How many people live in extreme poverty, on less than $1.25 a day?
We can't be sure.

Current estimate

350 million people may be uncounted – which could increase global poverty numbers by 25%

The data gap: we can't easily access enough information about how poverty is falling or rising in sub-Saharan Africa

31 countries have data on income poverty trends in the 2000s

13 do not

We need to find the missing millions. **#data2015**

developmentprogress.org

Source: WHO, UNICEF, UNFPA, World Bank and United Nations Population Division (2014), Carr-Hill (2013), PovcalNet.org

Development Progress

Introduction

About this paper

This paper starts by making the case for why a data revolution is needed, highlighting the parlous case of data globally, nationally and locally, focusing on social data, and stating why the gaps matter and how they should be plugged (Sections 1 to 3). Next, it discusses the innovative ways in which data are already being produced and used (Section 4). Finally we set out a vision for a fully-fledged revolution, with an examination of outstanding challenges (Sections 5 and 6).

Throughout, the paper highlights examples of a range of different uses for big and small data that are being piloted around the world. Some projects are already improving people's lives, while others are still works in progress. But they demonstrate that the data revolution is already underway.

Key terms

Household surveys

A household survey uses a questionnaire to gather information on households and their members from a sample of the population of interest, carefully selected so that the findings will be representative of the whole of that population (but with margins of error).

The major internationally comparable household surveys are the Macro International's Demographic Health Survey, UNICEF's Multiple Indicator Cluster Survey and the World Bank's Living Standards Measurement Surveys. These are usually conducted every 3-5 years. Most countries also implement national household surveys, which can be more or less frequent.

Census

A census is the complete enumeration of a population or groups at one time with respect to well-defined characteristics, for example: population, production, traffic on particular roads. Because it should cover the whole of a population, it can be used as a sampling frame for household surveys.

Administrative data

Administrative data come from one or more government sources such as hospital or school records. They are cheaper to collect than survey data, and should be more frequent and aspire to greater population coverage, but at present coverage in many countries is incomplete and quality of the data can be problematic.

An important type of administrative data is from civil registration systems – the governmental machinery set up in a country, state or province for legal recording of vital events (such as birth, death or marriage) of the population on a continuous basis.

1. What we do and don't know

Statisticians in Turkmenistan entering data into the database for further processing and analysis. Photo: © World Bank.

For the past 15 years, the international community has focused on a coordinated effort to transform the lives of poor people. The Millennium Development Goals (MDGs) were designed to revolutionise relations between donors and poor-country governments, galvanising action in developing countries, catalysing finance and other assistance where it was most needed, and celebrating improvements as and where they happened.

But, however worthwhile the aim, assessing whether the target has been achieved has too often depended on an illusion: that we really know the scale of the problems we are trying to solve, and that we can therefore tell what change looks like, and what has caused it. To put it starkly, we know a lot less about crucial facets of development than we think (see Box 1, page 15). We illustrate this here by showing data gaps and their impacts across several MDGs.

1.1 Data and the Millennium Development Goals

MDG 1: poverty figures and the missing millions

MDG 1 aimed to halve the number of people living in extreme poverty. But we do not know how many people in the world are poor. Attempts to measure consumption consistently and to arrive at an international definition of poverty are controversial (Chandy, 2013; Chandy and Kharas, 2014).

Sources of error are numerous, both in how the data are collected and when the World Bank aggregates these numbers to derive global poverty numbers (see Chandy, 2013). The problems are more acute for the poorest countries, where surveys are less frequent, and therefore estimates are necessarily based on a number of assumptions. For instance, just 28 of 49 countries in sub-Saharan Africa had reported household survey data on income or consumption between 2006 and 2013 as of that latter year – meaning, at the time 'a quarter of the 414 million people

who (were) estimated to live on under US$1.25 a day in the region according to the most recent official poverty estimate' were derived from earlier surveys (Chandy 2013, p. 14). For instance, in the case of Algeria, current poverty estimates are based on the 1995 household survey.[10]

The aggregation of country poverty into a global poverty figure introduces additional complications. One illustration of how these issues matter is in the dramatic revisions to global poverty numbers whenever a new global estimate is made – every 3 to 4 years. For example, in 2005, the estimation of 931 million poor people was revised upwards by a third to 1.4 billion, following revisions to price estimates and to the poverty line to account for inflation (Chen and Ravallion, 2010). Now the most up-to-date estimated number of people living below $1.25 a day is 1.01 billion, for 2011, nearly four years out of date, and there will no doubt be big changes (likely, reductions) this October, once a new set of figures are released based on updated data on prices (see Dykstra et al., 2014; Chandy and Kharas, 2014).

There is yet another complication: in poor countries, assessment of progress towards development goals is based primarily on household surveys, which generally omit *by design* the homeless, people in institutions, and mobile, nomadic or pastoralist populations (Carr-Hill, 2013). *In practice*, household surveys also typically under-represent people living in urban slums (because of the difficulty of identifying and interviewing), dangerous places and fragile or transient households. In other words, the very people that the Sustainable Development Goals are aiming to reach – with their emphasis on leaving no one behind – literally don't count. As many as 350 million people could be missed worldwide from these surveys and from many censuses according to work by the health economist Carr-Hill. He estimates that 250 million people are in groups the surveys are not designed to cover, such as migrants or pastoralists, while an additional 100 million may be under-represented because they live in areas which are difficult to survey. Considering the demographic they represent, it is likely that a substantial number of the estimated 350 million will be living on less than $1.25 a day**. In other words poverty figures could be understated by at least one-quarter** (Carr-Hill, 2013; PovcalNet.org). As Carr-Hill says, this systemic void makes a mockery of monitoring development progress because neither the baseline nor the current estimates are secure.

MDG 1 also aimed at eradicating hunger, but these numbers too have been subject to recent scrutiny. Notably, in 2012, the FAO revised its claim that the number of undernourished people in the world had climbed to over 1 billion in 2009, following food-price hikes, to suggest instead a figure of around 870 million. In other words, the figure was reduced by about 130 million (*The Economist*, 2012a).

MDG 2: universal primary education – probably overstated

UNESCO tracks data provided by schools to education ministries and collates the information into an annual report, the *Global Monitoring Report* (GMR). But this is very problematic as UNESCO itself openly states: official enrolment data may overstate the numbers of children in school at the appropriate age, suggesting that more needs to be done to address the problems of late entry and dropout. Household survey data for a number of countries indicate overestimates of 10% or more in school attendance rates. According to the GMR, administrative data may also be misreported and institutional incentives may play a role in this: if the number of students in the appropriate grade for their age determines the allocation of grants or teachers, schools and local governments might have a tendency to inflate the school register (UNESCO, 2010 and Sandefur and Glassman, 2014).

MDG 4: child mortality – uncertain

Child mortality, MDG 4, is regarded as the goal on which the data are best (Lopez, 2007). Indeed with the exception of the Pacific islands, data on child mortality appear to be widely available. Of the 161 developing countries, 136 have data to track this metric (Rodriguez-Takeuchi, 2014). Yet, according to data from the UN Interagency Group for Child Mortality, over two-thirds of the 75 countries which account for more than 95% of all maternal, newborn and child deaths do not have registries of births and deaths. And more than one-third do not have a child mortality estimate less than five years old. Twenty-six countries have no data on child mortality since 2009. As a result, child mortality data are estimates, often derived from surveys. In some cases these surveys collect very limited information on births and deaths, rather than full birth histories, so estimates are based on a number of assumptions (UN, 1983), such as fertility models that are unlikely to be representative of regions including sub-Saharan Africa (UN, 1982).

To see how this matters, take the example of Nigeria in 1998, where four different surveys of under-five mortality levels generated rates of death per thousand live births of: 225,[11] 186,[12] 173,[13] and 157.[14] The official UN interagency

10 PovcalNet.org (accessed 21 May 2015).

11 Using 2003 DHS and a full-birth history method.

12 2008 DHS, full birth history.

13 Malaria Indicator Survey.

14 2011 MICS and indirect method.

group modelled estimate for 1998 is 198 – somewhere in the middle – but the confidence intervals are sizeable.[15]

MDG 5: maternal mortality – a veil of ignorance

Most people, on reading a figure for maternal mortality, think this is the actual number of women who have died, expressed with forensic accuracy. Indeed, this is how the data are used: every year countries' performances are tracked against maternal mortality indicators. Results are published, and league tables drawn up with the achievement of one country judged against that of its neighbours, and other countries around the world.[16] Yet, the reality is that we do not know how many mothers are dying each year in poor countries, or even over longer intervals (Ronsmans and Graham, 2006).

How can this be possible? First, we lack adequate data. More than 100 countries do not have functioning systems to register births or deaths (World Bank/WHO, 2014), meaning that fewer than one in five births occur in countries with complete civil registration systems, with the remainder occurring either in countries with incomplete registries or with no data at all (2% of births – but 11 countries). Theoretically, the data could be taken from hospital registries, but many women in developing countries die outside hospital (PBR, 2007). This leaves household surveys.

Here too there are problems. The sensitivity of maternal death could lead family members to under-report it (PBR, 2007). More generally, maternal mortality is relatively rare – which is why it is measured per hundred thousand live births, compared to child mortality ratios which are measured per thousand live births. A typical household survey would find few respondents that have experienced a maternal death within their household or have a sibling who died of pregnancy-related causes.[17] It follows that the margin of error attached to these estimates is very large.

For most countries, therefore, maternal mortality is estimated based on a regression model that includes just three key predictors: per capita GDP, the general fertility rate and skilled birth attendance.[18] As a result of this modelling, we conclude that globally, for 2013, there were an estimated 210 maternal deaths per 100,000 live births. However, the uncertainty is such that the actual number could reasonably be (with 95% certainty) anywhere between 160 and 290 maternal deaths per 100,000 live births.

A maternal mortality rate (MMR) of 210 deaths per 100,000 live births translates into about 289,000 deaths. It follows that the difference between a MMR of 290 and a MMR of 160 is in the order of 180,000 deaths. In other words, there may have been 220,000 maternal deaths in 2013, or as many as 400,000.

If we look at just at the data for sub-Saharan Africa, the MMR is 510 – this means that about 179,000 women died in childbirth in the region in 2013, such that a woman faced a 1 in 38 chance of dying from causes related to pregnancy and childbirth in her lifetime. However, the uncertainty around this estimate is such that a reasonable MMR estimate could be anywhere between 380 and 730, which equates to a potential difference of about 123,000 lives. In other words, about 133,000 women may have died or about 256,000 (an almost 100% difference) – we cannot predict the number with greater precision.[19]

At country level, there is again considerable variation. In Nigeria, the maternal mortality ratio in 2013 was estimated at 560, but the confidence interval suggests the value may be anywhere between 300 and 1,000 (Figure 1). If the true number of maternal deaths were at the bottom end of the confidence interval in 1990 but at the higher end in 2013, then the number of maternal deaths may have risen over this period (Figure 1, overleaf) – the large margins of error put in question not only levels but trends (Melamed, 2014).

MDG 6: combating HIV/AIDS, malaria and other diseases

HIV infection rates have often been extrapolated from numbers among particular population groups for which such testing is mandatory – e.g., pregnant women visiting (often urban) clinics – which may be non-representative. In recent years, Ethiopia, Kenya and India reduced their estimated prevalence rates by half, following national population surveys (Kresge, 2007), and *The Lancet* recently reported a new methodology for modelling deaths from HIV which concluded that nearly 20% fewer people were living with the virus than previously thought (Murray et al., 2014). The data on malaria must also be interpreted with caution – we cannot be certain either about the incidence of malaria or trends in tackling it in countries

15 Computed from data at www.childmortality.org. The true figure could reasonably have been anywhere between 186 and 210 deaths per 1,000 live births, a potential difference of 132,316 deaths.

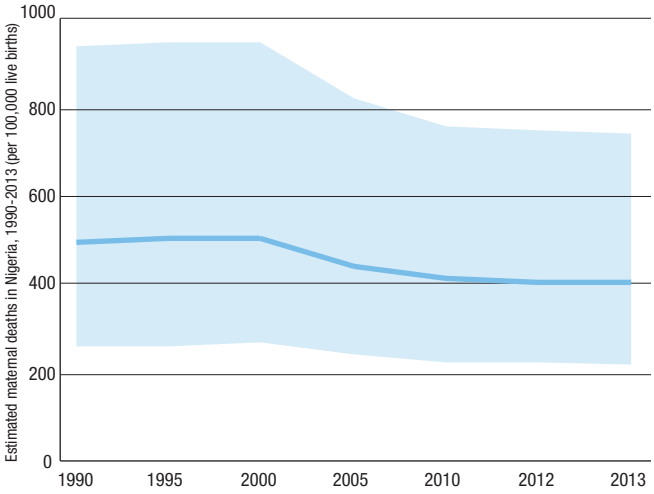
16 Multi-sectoral reports include the Global Monitoring Report, produced annually by the World Bank and IMF, and the World Health Statistics Report by the World Health Organization.

17 Maternal mortality is generally covered in household surveys by asking respondents if any women in their household died in childbirth within a certain recall period, or if any of their sisters died during pregnancy, birth or within two months of the end of the pregnancy (www.maternalmortalitydata.org/definitions.html).

18 The observed data on maternal deaths are used to validate these predictions. See Wilmoth et al., 2012.

19 In this section, numbers are taken from or computed on the basis of data in WHO, UNICEF, UNFPA, World Bank and United Nations Population Division (2014).

Figure 1: Trends in maternal mortality, Nigeria, 1990 to 2013



Note: The trend line depicts estimated maternal mortality and the shaded area shows the margin of error at a 95% level of confidence. Source: WHO, UNICEF, UNFPA, The World Bank and United Nations Population Division (2014).

that account for 85% of all (estimated) deaths from the disease. This leaves the claim that malaria deaths fell by 49% in sub-Saharan Africa between 2000 and 2013 open to considerable question (Melamed, 2014).

1.2 The people left behind

Most of these data gaps arise because current data collection techniques and sources are inadequate. One gap is temporal: internationally comparable household surveys are carried out every 3-5 years, meaning we have inadequate information between those years. The other key gap is in subject: many people (Carr-Hill, 2013) and many issues of great importance to poor people (Alkire, 2009) are left out. Importantly, for the SDG mission to ensure that the poorest and most vulnerable benefit from development initiatives, there are several discriminated-against populations about whom we know very little. These include women, persons with disabilities and those who are mentally ill.[20]

There are particular challenges in the collection of disaggregated data needed to identify the most marginalised. For example, the sample for the DHS 2011 from Nepal consisted of 11,000 households containing over 47,000 people. But suppose you want to monitor malnutrition among children under five from rural areas in the Far-Western region – a not unreasonable request for a national policy-maker: the sample size falls to 751 individuals. To then track performance of girls leaves a sub-sample of under 400 – enough to be statistically representative, but with a relatively large error. And if you want to add any additional filter, say smaller geographical regions or socioeconomic groups, the samples dwindle to such small sizes that it becomes very challenging to make any reliable inferences (Samman and Roche, 2014).

Increasing sample sizes may be possible, in some cases. But the larger the sample, the more expensive and difficult it becomes to get high-quality data. Pooling data across time is one possibility – but still problematic, in any country. It was estimated to take at least eight years of survey data to obtain reportable estimates for some population subgroups in the US National Health Interview Survey. An alternative is to use census data – but these are produced only every decade, and therefore inadequate for monitoring (Samman and Roche, 2014). Two marginalised groups least covered in terms of data are older people and ethnic minorities.

1.2.1 Older people

Although these numbers should be treated with caution, it is estimated that people over the age of 60 currently make up an estimated 12.2% of the world population, rising to

Box 1: 10 basic facts about development we still don't know*

1. How many people live in cities.
2. The volume of global assets which are held offshore, undeclared to tax authorities.**
3. How many girls are married before the age of 18.
4. The ethnicity of most Europeans.***
5. The percentage of the world's poor that are women.
6. Basic educational outcomes at primary level in sub-Saharan Africa, South-East Asia, Latin America.
7. The number of street children worldwide.
8. How many people in the world are hungry.
9. The size of sub-Saharan Africa's economy.
10. How many people work in the informal economy.

Sources:
Lucci, 2014; Buvinic et al., 2014; Marcoux, 1998; Klasen, 2004; UNESCO, 2014; de Benitez, 2014; The Economist, 2012b; Blas, 2014; La Porta and Shleifer, 2014.
Notes:
** There are data on these issues, but they are inadequate to give even a sufficiently accurate estimation.*
*** Zucman (2014); this figure could be between $8tn and $32tn (Henry, 2012).*
**** This is a controversial issue. See Open Society Foundations (2014).*

20 For gender gaps, see World Bank Group (2014). For gaps in coverage of persons with disabilities and mentally ill people see Samman, E. and Rodriguez-Takeuchi, L. (2013).

Box 2: DHS, LSMS and MICS: the three big household surveys

The vast majority of internationally comparable household data comes from three sources: the Demographic Health Survey (DHS), funded by USAID; the Multiple Indicator Cluster Survey (MICS), housed in UNICEF; and the Living Standards Measurement Study (LSMS), overseen by the World Bank.

DHSs are nationally representative household surveys that provide data on population, health and nutrition. These are collected through a household questionnaire, a woman's questionnaire (asked of all women in the household of reproductive age, usually 15-49 years), and a men's questionnaire asked of men aged 15-59. The standard survey is carried out every five years, with large sample sizes (5,000-30,000 households), and an interim one carried out in between, with fewer indicators and a smaller sample size.

MICS This includes questions on health, education, child protection and HIV/AIDS. It is of a similar design to the DHS, allowing comparisons to be made between the two. The surveys are organised into four questionnaires: households, women, men and children (with questions answered by the carer of children under five). MICSs have been conducted in 1995, 2000, 2005-06, 2009-11 and 2012-14. Survey sizes range between 5,000 and 20,000 households (Carr-Hill, 2013).

LSMS The World Bank's LSMS programme works with governments to improve their collection of household survey data. Questionnaires are administered to households, individuals and communities. The samples are nationally representative, but fairly small (2,000-5,000 households), meaning that accurate estimates can be obtained for the country and large sub-areas (such as urban/rural), but not at state or provincial level. These surveys are conducted every three to five years (Alkire and Samman, 2014).

16.3% by 2030 (HAI, 2014; UNDESA, 2013). Globally, there are already more older people than children under the age of five; by 2030 they will outnumber those aged 10 and under (UNDESA, 2013).

Aging brings with it a specific set of needs and challenges: households headed by older people, or with older members, tend to be poorer than others. At present, an estimated 340 million older people are living without any secure income and, if current trends continue, this number will rise to 1.2 billion by 2050 (Samman and Rodriguez-Takeuchi, 2013).

However, the DHS and the MICS currently ask detailed questions of women only up to the age of 49 in most countries,[21] meaning the circumstances and needs of older people cannot be fully assessed.

The INGO HelpAge International has attempted to fill this gap by producing a global index of wellbeing for older people. This is an important contribution, accounting for 90% of the world's older people (HAI, 2014) but it has suffered from data gaps. Only 96 countries are listed because those were the only ones with data available on all four elements of income security, health status, capability and enabling environment. Nor was it possible to identify differences between women and men, as the data were not disaggregated in this way.

1.2.2 Ethnic minorities

Being a member of an ethnic minority can be a key factor in passing poverty between generations (Bird, 2007). Ethnic minorities often face the biggest barriers to good health, and therefore have the worst health outcomes. Language and ethnicity may lead to marginalisation in education through complex channels (UNESCO, 2010). So monitoring progress explicitly in these communities may be particularly important.

But again, data gaps mean that ethnic minorities are often uncounted. In DHS surveys there are reliable, consistent ethnic trend data (or a sufficient proxy) for only 16 of 90 countries (Lenhardt, 2015). In some cases, the information exists, but is not consistent across years, making it difficult to estimate and track the progress of groups over time.

In Sierra Leone, information on ethnicities is available in the 2005 MICS but not in the 2010 survey. In India, the 2005/06 DHS lists scheduled castes, scheduled tribes and other backward castes, while the 1995/96 survey does not include other backwards castes (which in 2005/06 accounted for 29% of the entire sample). But for India, one of the most ethnically diverse countries in the world, even three categories does not give a sufficiently disaggregated picture (Mukherjee, 2013).

21 Because it was designed to survey women of reproductive age.

Box 3: Real-world policy constraints: the ODI survey

To confirm some of the anecdotal evidence about the lack of good data in developing country ministries, ODI interviewed a series of policy-makers based in line ministries to understand how they viewed capacity constraints in their respective countries.*

The policy constraints of not having access to up-to-date or readily useable data came through strongly. The most serious of these, concerning food costs, was cited by a respondent based in an African trade ministry, reporting a lack of data resulting in **sub-optimal policy-making** relating to maize imports and exports. He said: 'Data on stock levels of the main staple crop in the country and production estimates are the biggest constraint at the moment. The export ban on maize would most likely be removed if we had reliable statistics.'

Others reported that **data gaps impeded the government's capacity to provide services effectively**. A respondent from an African education department said: 'There is no way to follow individual children through their school careers and into adulthood. It is not possible to measure the returns to education, nor to see what predicts future employment opportunities.'

Three respondents identified constraints in setting economic policy. One, based in a national statistics office reported that a lack of data **seriously constrains the country's ability to calculate an accurate Consumer Price Index (CPI).** The weights are based on a partial household income and expenditure survey conducted in 1964 (and then only in the capital and on a sample of under 200 households), and the price data are currently also only collected from the capital city.

Another said: 'There is lack of clarity around basic public administrative statistics (personnel, spending) and a lack of detail to the data gathered from payment claims and vouchers which often makes it **difficult to clearly understand how and where budget is spent, and to devise ways to better execute the budget.**'

There are also problems with **stability and continuity of data collection**, particularly in countries in conflict. One respondent said, 'given the context [post-conflict] a lot of data was lost during the war. This means that for a lot of variables annual time-series are pretty short. Usually, this data can be found somewhere, for instance some civil society organisations have better records. But within the government the institutional memory was, for a big part, destroyed and no one really knows where to find the relevant data.'

A respondent in the Pacific islands noted that while data availability is relatively high, she recently needed to check historical data on export figures for crops. **Export figures for key commodities are collected by the central bank, but are not available online,** and have to be requested. 'The figures are presented in "containers" (as opposed to tonnes or kg) and it took about three weeks and consultations with three separate people within the bank to establish what a "container" might be.'

One respondent stressed that the problems of data in her ministry are political. She said, 'The ministry of finance is well supplied with the necessary IT, technical assistance and staff to access and make use of the data and statistics that we use. There are some IT issues of systems not being compatible with each other, but these can even be resolved... A shortage of solutions is not the problem. Rather, there's a limited understanding of how the public sector and civil servants in [country x] and in many developing countries work with data and how data serves them which makes development actors think there is a capacity issue. Furthermore, any issues which exist are exacerbated by development actors' training and the hiring away of competent staff.'

One respondent said that his **government was planning to build a multi-million dollar airport, but basing the decisions on scant data:** 'There was no cost-benefit analysis, and few pieces of data to underlie one. There were passenger numbers but not good quality enough for forecasting, especially once dynamic factors, e.g. growth of the economy, were taken into account.'

** The unpublished survey canvassed the opinions of 11 policy-makers based in line ministries in Africa, South America and the Pacific. Findings have been anonymised at the request of contributors.*

2. Why we need to fill the gaps

Participatory climate-service and information training in Tanzania. Photo: © Cecilia Schubert / CCAFS.

2.1 Policy implications of better data

If these data were used only by donors to assess to whom they should allocate aid, or by countries to assess their international ranking, this illusory sense of certainty, while not desirable, might not be so problematic. However, **at the national level, greater granularity, accuracy and timeliness are essential.** Here, data gaps are not a matter of theoretical imperfection: they mean that **vital knowledge that could be used to improve lives is missing.**

There is a wide literature on evidence-based policy-making, discussed elsewhere.[22] Here, we note the key aspects of government that require good data:

- spotting emerging policy concerns
- informing programme design and policy choice
- forecasting
- monitoring policy implementation and evaluating impact.

Without data, certain programmes would be very difficult to deliver: an obvious example is targeted cash-transfer programmes. But it is also clear that data are necessary to provide services systematically. While governments and NGOs are able to deliver improvements without perfect data, quality data are essential to reach the most marginalised in a way that responds to the specific needs of their communities, and to track progress.

Data gaps mean that governments cannot plan accurately. Despite the relatively extensive nature of data in Kenya, education planning is hampered by the lack of accurate country-level profiles of school participation. Watkins and Alemayehu (2012) argue that 'any estimate of out-of school numbers in Kenya is subject to large margins of error related to divergent estimates for the denominator (the number of children) and the use of different indicators for the nominator (the number of children enrolled or attending

22 For more discussion of evidence-based policy, see Head (2008) and Shaxson (2005).

Table 1: Reported school attendance and enrolment (Kenya)

Source	Attendance/enrolment (percentage)	Out-of-school estimate (millions)
Census 2009*	77	1.9
DHS 2008*	79	1.8
Uwezo 2011*	87	1.2
National administrative data**	90	1.1

**Out-of-school population calculated using the Census 2009 Primary School Population (ages 6-13).*
*** As reported on the Global Monitoring Report 2011 (ages 6-11).*
Source: Watkins and Alemayehu (2012).

school)'. Table 1 illustrates differences between enrolment and out-of-school numbers, according to data source.

This also means that governments cannot allocate their budgets efficiently. The Tanzanian government commissioned ODI to support its policy of more equitably allocating Local Government Authority (LGA) staff and funds by analysing its current strategy for grant allocations. The task, however, was difficult. During the timeframe considered, some budget lines were reclassified, making it difficult to compare spending across years. There were also changes in jurisdictions – in 2012 four new districts[23] were created and several had boundaries changed. In addition, different budget collection and management systems are used, making it nearly impossible to harmonise budget allocations with expenditure data. The education and health sectors, which were the focus of the study, are those with the most comprehensive budget data, implying that problems encountered in analysing other sectors in depth may be even greater.

Such gaps mean that governments (as well as NGOS, the private sector and citizens themselves) are not able to fix key public-policy problems because they don't know where or what the problems are. That child labour is a serious and persistent concern in Bangladesh is evident from talking to NGOs that work there.[24] But getting a complete understanding of the problem across the country, so the government can address it systemically, is not currently possible. This is ironic as there are three separate surveys in the country that could measure the number of children in work – yet none of them is adequate to the task:

- The most recent **Labour Force Survey** (LFS, 2010) only covers people over 15.
- The **Bangladesh Integrated Household Survey** (2011-12) covers the 6-11 age-group but the number of observations on child labour for this group (71 boys and 41 girls) is too small to permit rigorous analysis. Also because the survey does not cover household chores, the data on gender roles are skewed (as more girls than boys are likely to carry out such chores).
- The 2011 **Demographic and Health Survey** again does not cover employment among the 6-15 age group. Moreover, while the data are nationally and regionally representative, they are not necessarily representative of identifiable areas with high concentrations of child labour.[25]

Data can catalyse government accountability, although of course other things are also needed (McGee and Gaventa, 2010). A healthy democracy requires that the citizen should have access to honest and reliable information on public issues (SCUK, 2008). This is something that the majority of governments care about, even if it means different things in different contexts. Many will want to provide education, healthcare and a safety net to their citizens in the best way possible, and data can help by informing them in detail of needs.[26]

While gaps in data on marginalised groups mean governments (and civil society, academics and others) cannot reach far into societies and communities to find out about the most marginalised, these gaps also matter to citizens who need data to assess their own lives and communities. Absent this, they cannot assess whether budgets are being correctly and fairly allocated and whether they are missing out on opportunities to improve their own livelihoods. In fact, many people may not focus on the national level: the kind of power that matters may not be not wielded by national government, but by local government structures. It is at this level of disaggregation that vast improvements in data availability are necessary (Development Initiatives, 2014).

Quality data are essential to reach the most marginalised in a way that responds to the specific needs of their communities, and to track progress.

23 Usetha District Council, Msalala District Council, Kigoma District Council and Uvinza District Council.

24 See for instance, Bangladesh Labour Welfare Foundation (http://www.blf-bd.org/index.php/cms/priorities/#s).

25 Based on the 2005-2006 Annual Labour Force Survey. Understanding Children's Work Programme (2011) *Understanding Children's Work in Bangladesh*. Rome.

26 For an analysis of how this can work, see Fung et al. (2010).

Box 4: The value of data for evaluation

When Julio Frenk took office as minister for health in Mexico in 2000, more than half the nation's health expenditures were being paid out of pocket, and each year 4 million families were shattered by catastrophic health expenditures. Mexico ranked 144 in terms of 'fairness of financial contributions' in a 2000 WHO report. Frenk introduced the Seguro Popular, a public insurance scheme that was one of the largest health reforms in any country over the last two decades. A requirement to evaluate this – projected to cost 1% of the GDP of the twelfth-largest economy in the world in 2002 – was enshrined in law. Gary King, a Harvard data academic, was hired to carry out the evaluation. His experimental design paired communities with similar demographic characteristics, with one in each pair receiving encouragement to sign up for the insurance. The matched-pair design enabled pairs to be removed if needed (for example, owing to political intervention), without compromising the evaluation.

Subsequent analysis found that the programme was linked to a 23% reduction in catastrophic health spending among participant households, though the effect on health outcomes was not significant.

Sources: King, et al., 2009; King, 2014.

2.2 The economic value of better data

Good quality data yield not only social benefits, but also real economic returns, such that, in the medium term, a data revolution could pay for itself. First, if governments invest in better economic data, this can improve investor confidence. The IMF has found that, if countries invest in better-quality data, it is cheaper for them to borrow internationally. It investigated the effect of its data standards on sovereign borrowing costs in 26 emerging-market and developing countries and estimated that countries that sign up to its more stringent data standard reduce borrowing spreads (that is, the cost of borrowing) by an average of 20%.

Improving the quality of economic statistics can have other benefits too. In September 2013, the Kenyan government announced that the previous year's GDP was $53.4 billion – 25% higher than previously stated, and a level that now ranks it as a middle-income country, after updating the base year for its calculation. Growth for 2013 was revised up to 5.7% (Miriri and Blair, 2014). The IMF says that the rebasing made it easier to negotiate a new loan with the country (IMF, 2014).

Beyond this, other evidence is emerging on the value of data in other sectors. Many of these examples show the value to the private sector of releasing government-held data, allowing companies to be more efficient and innovative, although some also speak to the lower cost of doing government business with better information:

- McKinsey Global Institute puts the global value of better data and more open data at $3tn a year (with most of that accruing to the US and Europe) (McKinsey Global Institute, 2013). The main benefits would be seen in education, transport, consumer products, electricity, oil and gas, healthcare, and consumer finance.
- A report produced for the UK's Department for Business, Innovation and Skills estimates the economic value of the data held by the public sector which is then used by the private sector at £5bn a year (Department of Business, Innovation and Skills (UK) and Ordnance Survey, 2014).
- A report by PWC Australia commissioned by Google Australia found that, in 2013, data-driven innovation added an estimated AUS$67bn in new value to the Australian economy, or 4.4% of GDP, broadly equivalent to the retail sector's contribution (PWC, 2014).
- Corporate holdings of data and other intangible assets such as patents, trademarks and copyrights, could be worth more than $8tn, according to Leonard Nakamura, an economist at the Federal Reserve Bank of Philadelphia (Monga, 2014).
- A report by the World Bank, identifies several economic benefits of open data, including reducing the costs of existing services, and supporting creation of new businesses, digital service and innovation (Stott, 2014). Kenya, which in 2011 became the first sub-Saharan African nation to launch an open data initiative, estimates that opening up government procurement data and exposing price differences could save the government $1bn annually (Berkowitz and Paradise, 2011).
- In New Zealand, the government did a randomised controlled test on employment services. Without targeting, the service delivered NZ$1.20 (US$0.91) in saved benefits to a spend of NZ$1 (US$0.76). But when they used data on who uses the services most heavily – non-immigrant, young ex-criminals – to target provision, that figure jumped to NZ$4 (US$3.05).[27]

Better data can also result in important cross-departmental cost-savings for governments. Alex Pentland, co-creator of the MIT Media Lab says: 'When I worked in India, the head of the All India Institute of Medical Science told me that 90% of drugs were wasted because they couldn't track disease prevalence, so they had to hand out the same package to all districts.'

In the UK, following the release of data showing that some medical procedures were not effective, the Audit Commission found that the National Health Service could

27 Author's conversation with James Mansell, Member of New Zealand Data Futures Forum.

Box 5: Using new data techniques to save money on the census

The Dutch census of 2011 was carried out almost entirely by collecting data from registers, such as the new housing registry, in combination with other sources such as the land registry and the central bank. There were no household visitors to administer questionnaires. The census was conducted with a staff of 15 and a budget of €1.4m. By way of comparison, Germany – where register counts are combined with traditional census questionnaires – recruited 80,000 interviewers, and spent €750m on its census. Germany has nearly five times as many inhabitants, but even so the per capita costs were still far lower in the Netherlands. China's last census involved 6 million enumerators. Even with far lower wage costs, it cost €75m.

Of course, there are disadvantages to registry-based censuses. They are only reliable if based on high-quality central population registers; and setting up high-quality administrative registers, if they don't already exist, may be more expensive than carrying out a census. Presently, the coverage of registry and administrative data in many developing countries is incomplete and the quality of data is poor (Alkire and Samman, 2014).

Sources: Statistics Netherlands (2012) and (2014); UN Statistics Division (2013).

save £441m a year by not conducting them (Klein, and Maybin, 2012). The UK revenue authority (HMRC) is also using big data analysis of debit and credit card transactions to pursue tax evaders. No personal data identifying the card owners or their numbers is obtained by HMRC. Instead, the information is used to compare the amount of card sales a business makes each month with the taxes it pays – and action is taken if there are inconsistencies. One recent settlement netted HMRC more than £4m in unpaid tax (Ranson, 2015).

Professor Tim Hallett from Imperial College has shown that, in Africa, a given HIV budget prevents 16% more infections when targeted at the neighbourhoods where risk of HIV infection is largest than if the funds are spent across the entire adult population. Granular data can, in other words, dramatically reduce the cost of providing effective programmes (Over, 2014).

Moreover, the value of statistics and other data is long term – they have a long shelf life, and that value will only increase as existing data are used in new and different ways. For example, The Coalition for Evidence-Based Policy (2012) points out that the biggest cost of a randomised controlled trial is typically data collection; where administrative data already collected for other purposes, such as student test scores, criminal arrests, healthcare expenditures, can be used, costs can be dramatically reduced (cited in Alkire and Samman, 2014).

Similarly, in Colombia, the impact evaluation of a key conditional cash-transfer programme, *Familias en Acción*, was carried out using a combination of household surveys, a census of the poor, and the programme's own database. Commenting on this research, Rawlings (2013) observed: 'Linking all these data gave researchers answers in just six months at about one-fifth of the cost of an impact evaluation that that would require traditional primary data collection' (cited in Alkire and Samman, 2014; see Box 4, previous page).

2.3 Data as competitive advantage

New Zealand, a wealthy but geographically remote small island state, has been grappling with the issue of its competitive advantage since the 1970s when the creation of the European Economic Zone quotas meant its agricultural exports to Europe, formerly the mainstay of its economy, were significantly restricted. The country is now viewing *datafication* – that is quantifying things that might never before have been stored as data – of these same agricultural industries as the possible key to its economic future.

The government has recently endorsed the findings of the New Zealand Data Futures Forum, set up by the ministries of finance and statistics in 2013, and consisting of academics, civil servants, the private sector and Maori community leaders. The Forum identified 'high-value opportunities for New Zealand to position dairy, meat, wood, horticulture, fishing, wine, forestry, wool and tourism in the emerging data future. Datafying these industries, and creating strategic data assets to underpin innovation and decision-making, can give us a competitive advantage' (NZDFF, 2014).

Without making the same kind of wholesale commitment, the UK has also identified big data as one of eight areas of technology in which the country has a competitive advantage (Willetts, 2013). Poor-country economies may differ in significant ways from G20 ones. However, it seems likely that the gains in developing countries from improved and new forms of data could be even greater, as these countries start from a lower base (Giugale, 2012).

3. How to plug the gaps

A worker collects data from women in Badakhshan province, Afghanistan, as part of a village-level assessment of hazards, vulnerability and risk. Photo: © Luke Bostian / Aga Khan Foundation.

If the data revolution is to ensure that these gaps are plugged, it will need to strengthen NSOs and improve the use of existing data. These two elements are discussed in this section, while Section 4 considers potential new sources of data.

3.1 Strengthening NSOs

3.1.1 Lack of capacity

For years, NSOs in developing countries have been under-resourced, and this has had clear implications for their ability to collect quality data (PARIS21, 2015). In addition, populations in developing countries may live either highly spread out or in complex, dense and ever-shifting communities. This means that statisticians, asked by the global community or their government to come up with finite numbers, have often had to rely on long series of assumptions and interpolations. In other words, they do the best they can with the limited human and financial resources available to them.

Data collection is expensive. Increasing the range and depth of questions and number of interviewees to improve the quality of the data adds to that expense, and can compromise data quality when survey fatigue sets in. Already the India National Sample Survey Consumption Poverty module takes between 1 and 2.5 hours to complete – and this gives data on only consumption. If benefits are not obvious, it may be difficult for line ministries to justify spending on surveys rather than investing in say, health systems, where outcomes may be more tangible. Therefore while it is clear that investment in NSOs is needed, new methodologies and technologies for data collection and communication will be also be necessary if countries are to significantly improve the quality of their statistics.

3.1.2 Political interference

Data are inherently political, and may be more a result of negotiation than measurement. **In particular, the decision as to what (and whom) to count is intimately connected with questions of power** (see Boxes 7 & 8, overleaf). In other words, groups that are marginalised may be intentionally so (Buettner and Garland, 2008 cited in Carr-Hill, 2013; UNESCO, 2010). Both the MICS and DHS provide governments with the option to include questions on ethnicity, for example, or to use proxy measures such as religion or language. In MICS most governments have chosen not to include the questions or the proxies (Dodgeon, 2013).

Box 6: National Statistics Offices (NSOs)

In a PARIS21 survey of 25 developing country NSOs, only half said that they follow international standards, and a fifth have no quality control in place. In the Dominican Republic, Nepal, Trinidad and Tobago, and Bolivia, laws that govern statistics are more than 35 years old. The same study found that the NSO in Burundi doesn't even have an office.

A respondent to the ODI data survey outlined in Box 3 who is based in an African statistics service said that eight out of nine people in the office did not know how to use their computer mouse, let alone MS Excel. Another reported that staff in the statistical ministry had 'minimal to zero ability to analyse or use data'. Typically data are 'outsourced to third parties for analysis'. Another said that no one other than himself knew how to use STATA or SPSS (computer programmes to analyse data).

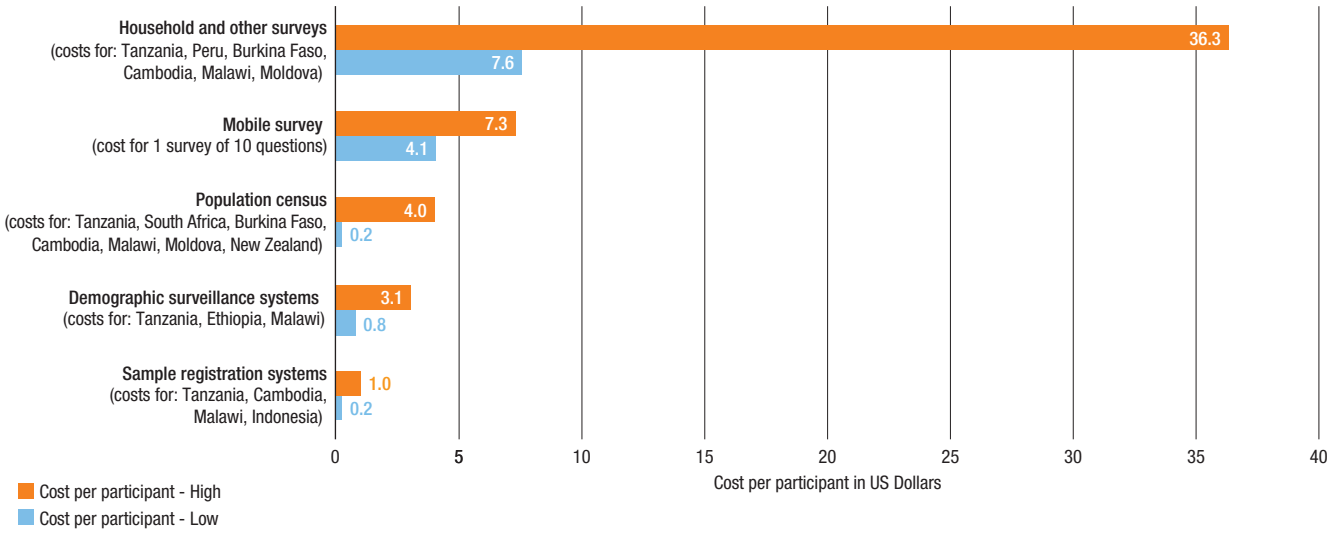
In NSOs in Malawi, Mali, and Trinidad and Tobago, fewer than 20% of staff members have any training in statistics, and there is no training plan in place (PARIS21, 2008-2015). Low pay is also endemic. One respondent to the ODI survey said that the World Bank paid for training of four statisticians in the ministry of finance but 'a couple of years later, only one of them still works in the ministry. The rest moved to the private sector, where they gain more. The last one is also looking to move.'

Despite of these manifold challenges, some offices are nonetheless innovating. Costa Rica, Gabon, Peru and the Philippines are using commercial data in income and expenditure surveys. Almost 50 African NSOs now have websites, which carry nationally representative surveys and censuses.* Some NSOs now use Twitter.**

** See for example Kenya National Bureau of Statistics (http://www.knbs.or.ke/).*

*** As of February 2015, among others the NSOs of Mali, Senegal, Burkina Faso, Kenya, Mozambique, Nigeria, Uganda, Rwanda and South Africa had what appeared to be an official Twitter account.*

Figure 2: Cost estimates for alternative data-collection approaches



Note: A DSS monitors demographic and health outcomes for the entire population of a given area. A SRS records the births and deaths within sample units throughout a country to estimate population, fertility, mortality and medical attention at birth and death.

Source: Health Metrics Network (n.d.); Croke et al., 2012 and authors' correspondence with GeoPoll.

In India, caste had not been included in the census since 1931, but was included in 2011 for political rather than ethical reasons: the ruling Congress party needed the support of smaller parties including caste-based groups (*The Economist*, 2010).[28] A Socio-Economic Caste Census (SECC) has also been conducted for the purpose of targeting, but has not yet been released.[29] For NSOs to provide accurate and complete data, they will need to be politically independent from government (see Box 6).

28 See also Census of India (2011) (www.censusindia.gov.in/).

29 Socio Economic and Caste Census 2011 (http://secc.gov.in/welcome).

Box 7: Counting Myanmar*

Myanmar has just conducted its first census in 30 years, and initial findings are dramatic: there are 9 million fewer people than everyone had thought: companies, donors and the government had long spoken of Myanmar as a country of 60 million people. Yet the census showed just 51.4 million – around 15% fewer. There are also nearly 2 million more females than males. Myanmar is a vast land with poor physical and digital infrastructure and dispersed populations. In Putao, northern Kachin state, enumerators walked for 12 days to count just a small collection of households. And after years of military rule, government activity is still viewed by many with the highest suspicion. The government is still technically at war with a number of ethnic-based separatist armies in its outlying states.

UNFPA, supporting the Myanmar Department of Population, rapidly trained more than 110,000 enumerators. To engender trust, they employed primary school teachers fluent in local languages and known by communities, and UNFPA's presence as a foreign source of expertise helped to allay suspicions around the intent of the census, securing the eventual agreement of 14 warring ethnic bodies to facilitate enumeration in their communities. Much of the estimated $62m cost (Jackson, 2014) has been spent on communication efforts to explain the value of the census. Around 7.5 million pamphlets in 19 languages were distributed, along with 2.3 million posters, radio broadcasts in eight languages, 500,000 t-shirts, 1 million census caps, badges and fans, and a 24-day celebrity bus tour.

In the end, 98% of households responded to 40 questions that will finally provide a firm statistical basis for public policy-making and donor interventions in Myanmar. However, the source of the missing 2% is contentious.** The question of how the oppressed Rohingya community would be classified dogged the process from the outset. On the day of the first enumeration the government declared that Rohingya people would not be allowed to self-identify. Where individuals identified as Rohingya (as opposed to the government's preferred classification using a Bengali term implying belonging in Bangladesh), many enumerators simply closed their manuals and moved on to the next house.

This caused uproar among many who felt they were, perhaps intentionally, being left out of the census (Snaing, 2014) and widespread international condemnation of a wider campaign to delegitimise Rohingya people and deny them citizenship in Myanmar (Mahtani and Myo, 2014). Modelling work was done to account for these missing households, but human rights groups have criticised both UNFPA and the government. It remains to be seen whether this will undermine the findings once they are officially published.**

** For a critical review of the census, see The Economist (2014b).*

*** Missing Rohingya households do not constitute the entirety of the uncounted households. A small region within Karen state, entirely consisting of local independence movement fighters, could not be visited as soldiers were unwilling to disclose personal details. Some households were also unable to be reached in the most remote locations.*

**** Author's conversations with Janet Jackson (Myanmar Country Representative) and Petra Righetti (Census Coordinator, UNFPA).*

3.1.3 Perverse incentives from donors

Donor behaviour can undermine coordinated country-led data collection. Frustrated by the lack of domestically generated data, donors (and NGOs) commission surveys that answer their monitoring and evaluation questions at a time that works for their budget cycle. This can come at the expense of domestic statistical capacity, as staff from NSOs are frequently hired to do the surveying on a consultancy basis, hence undermining the already-stretched capacity of the NSOs.

Donor-sponsored surveys are not always sufficiently coordinated, either with one another, or with country surveys. For example, in Mozambique the Household Budget Survey (*the Inquérito ao Orçamento Familiar*, which was the LSMS equivalent)[30] and the MICs (*Instituto Nacional Des Estatisticas*, 2009) survey both happened in 2008/09, but there had been no household survey in the previous five years. This means that countries have important – and unnecessary – data gaps. This chimes with a comment made by a respondent to the ODI data survey (Box 3, page 17). Referring to the DHS and MICS, he said: 'It is hard to describe the amount of damage that donor-sponsored surveys are doing to statistical capacity. Due to the inflow of donor funds, the stats agency had almost zero interest in doing any work for the government – which didn't come with salary boosts, per diems, etc.'

The DHS and MICS are now trying to coordinate their work with LSMS and ensure that what they produce is comparable[31] (see Quick Win 2, Box 10, page 28). In addition, there has been an effort in recent years to align donor support to developing country national priorities through the National Strategies for the Development of Statistics (NSDS).[32]

30 See Mozambique Household Budget Survey 2008-2009 (http://ghdx.healthdata.org/record/mozambique-household-budget-survey-2008-2009).

31 Author's conversation with Gero Carletto, Chief Economist, LSMS, World Bank and Sunita Kishor.

32 See PARIS21, *National Strategies for the Development of Statistics* (http://www.paris21.org/NSDS).

Box 8: Pakistan: poverty reduction or political manipulation?

According to its most recent Economic Survey, Pakistan is on track to meet the MDG target for reducing extreme poverty. The proportion of people below the national poverty line dropped from 26.1% in 1990 to 17.2% in 2007, and to 12.4% in 2011.* Wonderful news – but is it true? It is certainly counterintuitive: between 2007 and 2011, the country experienced an economic downturn (albeit following a boom), high prices, a series of significant electricity cuts and major natural disasters.

The poverty numbers are subject to fierce debate in Pakistan, and some suspect manipulation of data for political ends. But even that may be too simplistic an explanation. Unpacking the real story is complex, and separating political from technical aspects even more so. The official estimate for 2007 of 17.2% was not believed by some parts of the government. An independent panel of economists used preliminary survey data from that year to show that 35-40% of the population was living below the poverty line (Naseem, 2012). But their methods were also questioned. In 2009, the World Bank was called in, verified the 17.2% figure, and published it in its Country Partnership Strategy, although Pakistan still refers to the figures as 'interim' (Naseem, 2012).

But there's more: according to Dr Talat Anwar, a former adviser to the Planning Commission, the government has deliberately underweighted the share of food in the consumer price basket since 2008, at a time of particularly high food prices, to under-report inflation. He says that this will have affected poverty estimates (Recorder Report, 2015). A technical group in the Planning Commission has now been asked to review official methodologies and determine possible causes of variance in the poverty numbers (Government of Pakistan, 2014). Until this is done, and until the Pakistan Bureau of Statistics operates independently of government, accusations of manipulation are likely to continue – and it still isn't clear how many people in the country are poor.

** According to the Pakistan Economic Survey, this number should serve as an interim indication of poverty. A technical group is revising the official methodology and updating the estimates (Government of Pakistan, 2014).*

Source: Khan et al., 2015.

Finally it may matter deeply to governments if data show that they are not increasing the number of children completing primary education, or that they are failing to reduce malnutrition. This not only affects their domestic political standing, but may also mean donors cut aid programmes or other funding. In other words, they have an incentive to report that things are better than they actually are (Sandefur and Glassman, 2014).

3.1.4 Making existing data transparent and usable

There are vast repositories of data in the world and one of the most cost-effective ways for countries to improve their statistical capacity is to use that data by publishing as soon after collection as possible. But beyond this basic point, there are key ways to improve the usability of data. The first concerns how the data are collected. To be genuinely revolutionary, **data will need to measure issues that citizens in poor countries care about**. An obvious way to ensure that the data are relevant is to have users engaged in data projects from the outset (NZDFF, 2014).

It is also essential that the users themselves directly benefit from the data collection (Bhargava et al., forthcoming). In the case of New Zealand, one explicit objective of the Data Futures project was to improve livelihoods, and help people get better access to services and benefits. As the report says: 'After all, Maori data is part of the Maori story, but is too often only analysed and interpreted by providers of services to Maori for their own purposes, not to answer questions put to those agencies by Maori' (NZDFF, 2014). The government is also exploring how to improve service provision to the Maori community based on insights from data.

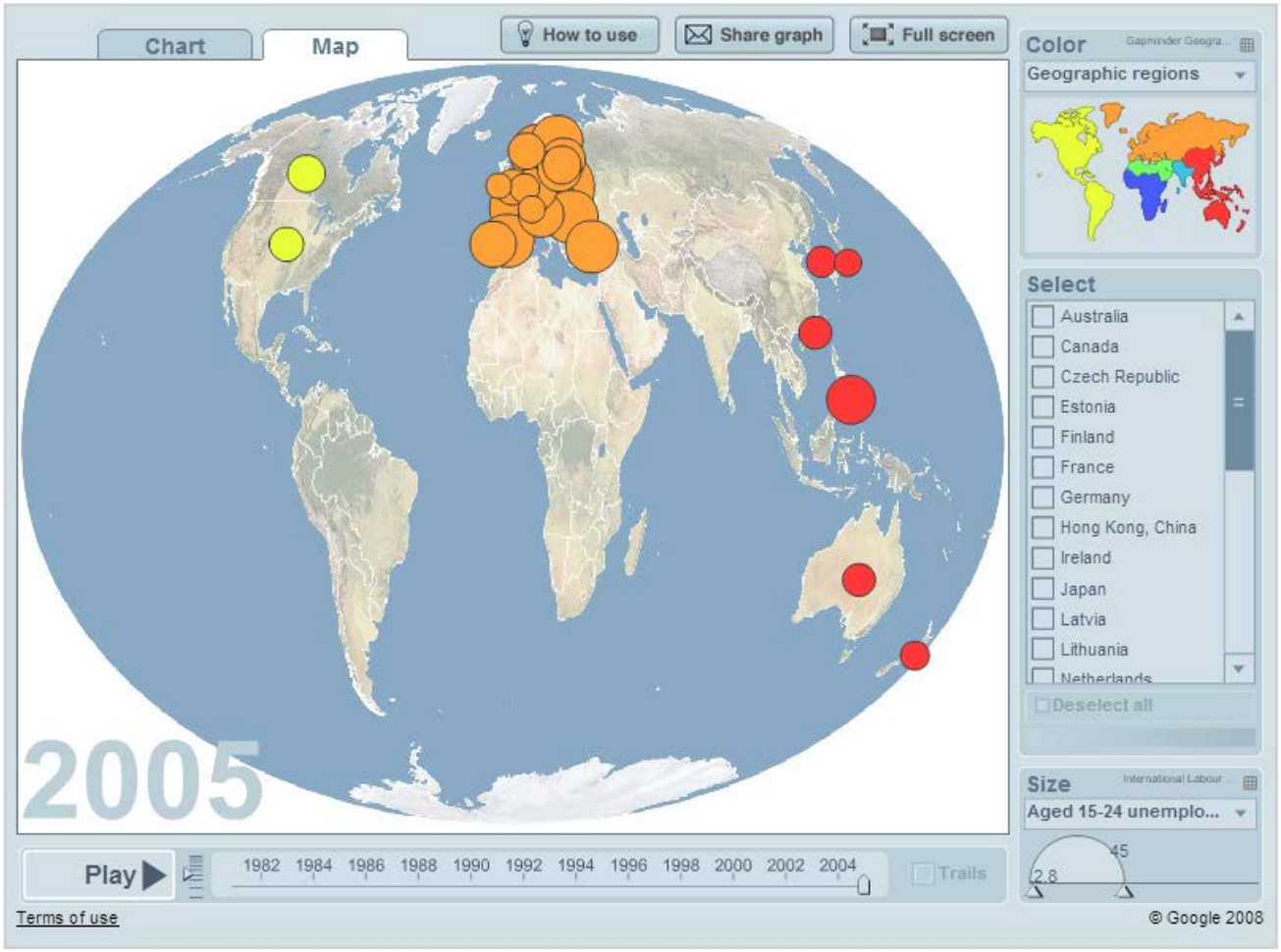
The data need to be useable. This might seem an obvious point, but making data accessible is not as simple as putting large volumes of numbers on a website. They need to be standardised so that they can be presented in a uniform way, and published in a format that people can understand. The World Bank boasts that its open-budget portal, the Boost Initiative, is user-friendly, but, ironically, it is for the most part impenetrable even to professional researchers.

The data need to be searchable, or machine-readable. In India, for instance, government documents are published online, but not in a digitised way – in fact, there is a kind of reverse digitisation: documents are created on a computer, then printed out, stamped with an official seal and signed in ink by an authorised person, scanned, converted to a PDF and then posted (Abraham, 2014).

Then **the data need to be understood**. There needs to be some basic capacity on the part of the user to understand the data, not least for ethical reasons (Bhargava et al., forthcoming). There have been some efforts to teach statistical literacy in developing countries, for instance the maths4stats initiative, which was trialled in KwaZuluNatal in South Africa (North et al., 2014).

But expecting that the broad population will understand data sufficiently well to have the capacity to interrogate them is a large assumption. Moves towards data visualisation are very helpful and can allow people to

Figure 3: Example of a Gapminder graphical representation (% population aged 15-24 registered unemployed)



Source: Gapminder.org.

analyse data who otherwise would not be able to (Figure 3). Hans Rosling's Gapminder Foundation, for instance, has created a way of teaching statistics that is vibrant and compelling.[33] But visualisations may still not be useful for the very poorly educated, may be less clear to some people than numerical data, and are themselves interpretations of data that may be subjective (Bhargava et al. forthcoming). In the immediate instance, data intermediaries will be needed (see Jooste, 2013). NGOs, foundations, and even political parties could play this role where the private sector does not.

Finally, in order **to serve development, it is essential that data reflect the complex adaptive nature of systemic problems**. We know, for instance, that enrolment is a poor metric for education, hence there needs to be a shift to measuring years in school or learning outcomes. We may never know the exact number of poor people, as this will constantly evolve: as soon as a survey is conducted, it is out of date. In the face of this uncertainty, the real potential of data is in providing real-time feedback. The aim is not a definitive number of maternal deaths but rather real-time updates on whether this value is going up or down, or whether women are reporting better or worse care during birth, for example. This requires changing how we view data: not as something static and fixed but as always evolving, and that helps us to understand and respond more quickly.

33 http://www.gapminder.org/for-teachers/.

Box 9: Input data

In much of the conversation about data to support implementation of the MDGs, the focus has been on results and outcomes. However, it is equally vital to track inputs: financing through budget revenue, aid and other development finance; and government spending in relevant areas. Without this, it will be impossible for stakeholders to hold governments and other actors accountable for funding delivery of the SDGs. Experience in this area suggests both optimism and room for improvement.

- **Improved data on inputs can improve results**. Where data on inputs have been more transparent and clearly linked to the MDGs, accountability processes, government planning and budgeting, and MDG outcomes have improved – as in Zambia, where the 'Vote Health' campaign used a pre-election period to demand a dramatic increase in health spending which government data showed was well below the agreed target (15% of spending) (IBP/DFI/Oxfam, 2014).*
- **The countries that provide readily accessible expenditure data by sector** are not necessarily the ones you would expect: Afghanistan, Burkina Faso, Cambodia, Guyana, Honduras, Mozambique, Solomon Islands and Uganda.**
- **Data on inputs are woefully insufficient in many countries**. Many countries still do not compile or publish budget data in ways that allow people outside (and inside) the government to track spending to achieve development goals, particularly in the water, sanitation, and environment sectors, and on gender. More spending information is available on education, agriculture and health.

** The link between transparency and MDG spending allocations is complex. Across the whole sample, more transparent countries do not on average allocate a higher share of the budget to education, health or water. However, indicative results appear stronger for countries that have seen recent improvements in transparency levels.*

*** Open Budget Survey, Data Explorer (survey.internationalbudget.org/#rankings).*

Box 10: Quick wins

Below is a set of proposals that could be achieved relatively easily, in terms of money, time and political will, and would be a significant fillip to the data revolution.

Quick win 1: Open up the World Bank's hidden database

Even though the World Bank is supposed to operate an open access policy (World Bank, 2013), it manages a secret database – secret in the sense that it is accessible only to World Bank staff (and consultants). **The International Income Distribution Database (I2D2) contains data from 600 nationally representative household surveys. This information would be hugely useful if made public** – it is the only database in the world that gives comparable fully disaggregated data on income poverty down to the level of individual people. The data need to be verified for some regions, but this would then be an extremely important source of disaggregated data. Although the data belong to country governments (because they come from their national household surveys), there is no reason to expect that they would not give permission for the data to be made publicly accessible.

Quick win 2: Coordinating the DHS, MICS, with the LSMS

The DHS and MICS are not always coordinated with the LSMS, meaning that their data are not always compatible. The three surveys should be better aligned so they occur at intervals rather than being bunched into the same year (which would result in a more regular flow of information); use the same sampling frame; and, for key indicators, word questions in similar ways.

Quick win 3: Modules on missing dimensions of poverty (Alkire, 2007)

Household surveys could include optional modules asking specific questions on dimensions that are rarely measured but that deprived people cite as important in their experiences of poverty. According to the Oxford Poverty and Human Development Initiative (OPHI), these include quality of work, empowerment, physical safety, social connectedness and psychological wellbeing. OPHI has designed short questionnaires (8-10 minutes) that could be integrated into household surveys to give insights into these so-called missing dimensions. These modules contain indicators selected to be internationally comparable and to measure change over time.

Quick win 4: Experiment with linking survey and administrative data (Alkire and Samman, 2014)

Experiments with linking survey and administrative data, primarily in Europe, have shown considerable potential to enrich our knowledge of households and individuals – at relatively low cost. Data may be linked between individuals and households from disparate sources, or between households and communities. For instance, in the 2009 Human Development Report for Mercosur (UNDP, 2009), a multi-dimensional poverty index for young people was constructed which included one dimension – health and environmental hazards – composed of indicators coming from administrative data at the state level in Brazil and at the provincial level in Argentina. These efforts are not free from challenges but the potential that linking offers to capitalise on the strengths of both household surveys and administrative registries warrants further investigation and experimentation.

Quick win 5: New technologies*

New innovations, notably Computer Assisted Field Editing (CAFE), Computer-Assisted Personal Interviewing (CAPI) and cloud-based technology, can facilitate collecting data and making them available quickly. CAFE permits entry and editing of survey data while in the field – resulting in swift data entry and the quick diagnosis of any problems. CAPI is an entirely digital method of data collection. It too permits the immediate transfer of data to central offices for ready analysis. A quick count of ongoing DHS surveys shows that about two-thirds are using either CAPI or are using CAFE. Other data collection methods using new technologies seek to involve the respondents more actively in both data collection and their analysis. For example Paraguay's Poverty Spotlight relies on a 20-minute visual survey methodology that enables people who are poor to create innovative maps showing the dimensions in which they are poor using stoplight colours (red, yellow, green), photographs, maps electronic tablets and simple software.

Quick win 6: Investing in data

At the Financing for Development Summit in July 2015, in Addis Ababa, governments should commit sufficient funds to finance a significant improvement in statistical systems. There are various estimates of what this would cost. Initial estimates from the Sustainable Development Solutions Network (SDSN) suggest $1bn annually to improve national statistical systems so they could measure the SDGs (Espey et al., 2015). Demombynes and Sandefur estimate the cost to international donors of improving household surveys would be $300m a year (Demombynes and Sandefur, 2014). In other words, costs are not likely to be substantial, and it is clear that these are investments that deliver a significant degree of value for money immediately and in the longer term.

Quick win 7: Continuous or interim surveys (Alkire, 2014)

To increase the frequency of data, some governments and international agencies have implemented continuous surveys or interim surveys with shorter questionnaires, smaller samples and, sometimes, rotating modules. For surveys to be continuous, different surveys are drawn from a master panel. These can be combined for more in-depth disaggregation and may have a panel element. Examples include national household surveys in Indonesia, Ecuador and Brazil, and DHS surveys in Peru and in Senegal. A key challenge for interim surveys is to generate enough demand for a short instrument. The Multidimensional Poverty Peer Network (MPPN) has proposed a 'short, powerful' interim survey to measure core SDG indicators of human poverty (and how they are related at a household level).

Quick win 8: Open by default

The International Household Survey Network and Accelerated Data Program, managed jointly by the World Bank and PARIS21, have supported many countries to establish data warehouses and to provide access to micro-data from surveys.** However, it is still the case that vast amounts of globally held data are unpublished. This not only in itself potentially a huge waste, it also means that surveying is often duplicative. Donors should make their engagement with international agencies conditional upon those agencies automatically making their data open access.

Quick win 9: Improving civil registration and vital statistics (CRVS)

Already some developing countries are moving towards a full registration system of births and deaths. This is incredibly important to citizens, who may need a birth certificate to claim property rights, access services such as education, and avoid becoming stateless in a disaster. It is also important to governments, who cannot readily track progress in poverty reduction and economic growth absent a complete picture of the population (World Bank/WHO, 2014). Latin America has been particularly successful in creating such a system. In Africa, success stories include South Africa and Cape Verde (World Bank/WHO, 2014), and large-scale efforts are now underway in the region.*** All countries should put in place quality CRVS systems. Pali Lehohla, the South African statistician general, says that, from a standing start, it should take countries just one year to develop a functioning system. He adds that CRVS systems are inexpensive: in South Africa, the cost has been US$0.20 per capita.†

Quick win 10: Keeping data systems open

As governments develop data infrastructure, they should invest in open-source systems†† rather than proprietary ones. This ensures that they retain ownership of the data, plus the possibility that they can migrate them in future. This also means that, once they have the technical capacity, countries can shape the system as they want, and resolve problems in-house rather than depending on experts outside the country.

* *Cited in Alkire and Samman (2014); Alkire (2014); Rojas (2015); and personal communication with Sunita Kishor.*

** *www.ihsn.org/home.*

*** *See for example, the Third Conference of African Ministers responsible for Civil Registration Yamoussoukro, Cote d'Ivoire 11 February 2015 (www.uneca.org/sites/default/files/uploaded-documents/CRVS/2015/paris21_concept_note_crvs_yamoussoukro_v04.pdf).*

† *Author's conversation with Pali Lehohla, Statistician General, Statistics South Africa.*

†† *Computer software with its source code made available with a licence in which the copyright holder provides the rights to study, change and distribute the software to anyone and for any purpose.*

Box 11: Innovations in data collection and mapping in Liberia – from water points to Ebola

As Liberia emerged from its civil war, the country's infrastructure was, not surprisingly, in disarray. The government faced the enormous challenge of prioritising investments and planning. One particular challenge was the huge number of people in rural areas without access to clean drinking water.

New technology has been improving the government's access to information: A country-wide exercise to check and map water points has given the ministry of planning a reliable database to guide its decision-making.* The government of Liberia has said that the water-point map, showing 10,000 pumps across the country, was the main input into the water and sanitation elements of its national plan, Vision 2030, particularly targeting underserved areas. 'It helped us to justify why those resources should go to one county as opposed to another,' said the then assistant minister for community services in the Ministry of Public Works, George Yarngo.

The 2011 survey took six months to complete. Enumerators interviewed communities and photographed the water points: the software used enables any Android phone to geo-locate, photograph and collate details of water points, which are uploaded to a central database. The entire, searchable dataset was put on a website (www.wash-liberia.org). It showed that just under 64% of wells and pumps were in good working order, 11% were functioning but had some kind of problem, and 25% were not working.**

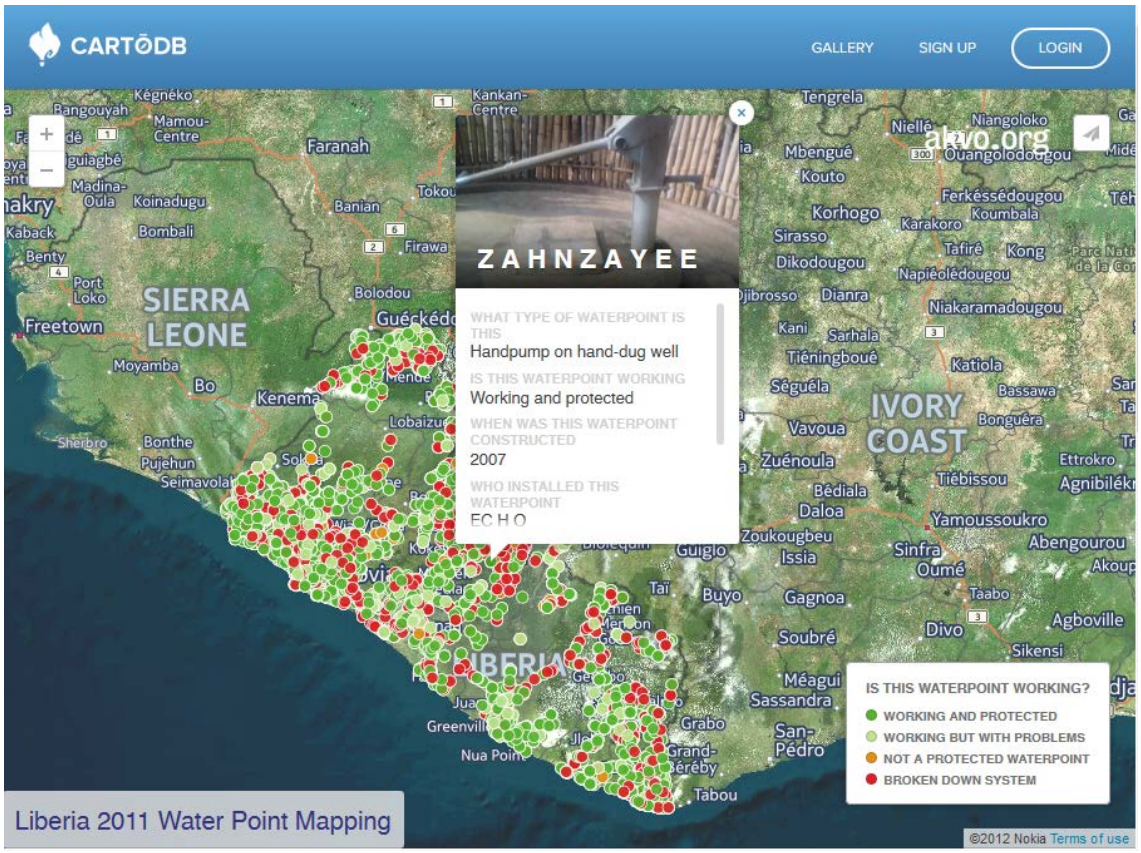
The government says that it intends to keep the map updated, but work was interrupted by the onset of the Ebola crisis. However because people were already trained in the software, they are now using it to collect information on health facilities and to survey people concerning the impact of the disease on livelihoods.*** Similar mapping using the same software is happening in other countries including Ghana, Ethiopia (including in the fragile Somali region) and Vanuatu.

* *http://akvo.org/blog/mapping-the-pipeline-liberia.*

** *http://wash-liberia.org/data-maps.*

*** *http://wash-liberia.org/ebola.*

Figure 4: Liberian water point mapping



Source: Liberia 2011 Water Point Mapping, Akvo.

4. A data revolution: already happening

Interview crowd in Daykundi, Afghanistan. Photo: © Institute for Money, Technology and Financial Inclusion.

Despite the gaps and deficiencies outlined above, many governments are already investing in data, improving their NSOs, and experimenting with new ways of producing and using data to deliver progress.

4.1 Examples of innovative approaches to data production

Costa Rica, Brazil, Mexico, Singapore, Indonesia, and the Philippines, as well as developed countries such as the US, Sweden and Japan, are already successfully innovating with and opening up data to solve complex policy problems, increase allocative efficiency and improve democratic processes. The US has just appointed its first ever Chief Data Scientist (Patil, 2015).

The entirety of **Costa Rica**'s procurement information is online, in raw form. On launch, researchers who worked with the data (that is, sorted them in an Excel file) found that 20% of all government purchases were going to 20 companies – which basically meant 20 families. This revolutionised government contracting: every bid is now public, and the entire process can be tracked online.[34]

Rwanda has dramatically improved its data production, particularly its core economic statistics. The Consumer Price Index, Producer Price Index, and Gross Domestic Product are now produced and published on time. The release of key statistics – such as those from the Demographic and Health Survey and Living Conditions Survey – were timed to allow assessment of progress against Rwanda's first poverty reduction strategy and inform planning for the next one (Glassman and Ottenhoff, 2014).

The government of **Mexico** is investing heavily in innovative means of capturing and using data. It has established census units in each state, and now quickly processes and releases data from household surveys. The country's lead institution on poverty measurement and

34 The system is called Mer-Link (http://www.mer-link.co.cr/index.jsp).

monitoring, CONEVAL, obtains data from the national household survey and posts official multidimensional poverty statistics – which include income poverty – nationally and by state two weeks after receiving the cleaned data, and in a way that allows researchers to verify them (Alkire and Samman, 2014). The NSO is fully independent of the government.

Mexico also uses an innovative method to count maternal mortality. According to Eduardo Clark, director of Data for Development in the Coordination of the National Digital Strategy, states used to have incentives to under-report such deaths because they were seen as evidence of poor governance. States now record 84 variables at the time of a child's birth, of which 69 are made available in Mexico's open data portal (datos.gob.mx). These high-quality data are then compared with data on death certificates, which are analysed by physicians at the Ministry of Health to ensure they are correct.[35] In 2011 the official death certificates originally identified 943 cases of maternal deaths for the year. After reviewing suspicious death certificates, the Health Ministry identified an additional 124 cases, representing a 13% increase.

The **UK** Office for National Statistics (ONS) is currently experimenting with web-scraping – that is, extracting data from web pages – as a possible additional source of data for price statistics. Traditionally, data have been collected by physically visiting shops to gather price information, but web-scraping allows extraction of price data from websites, meaning an index could be produced in near real-time and very cheaply. This is a pilot project focusing on supermarket groceries and collecting price quotes for products from around 40 item categories. A major early challenge is that a third of products change each year, making it difficult to track price changes for individual products and so ensure the index is comparing like with like. It is also not yet clear what the demand would be for such high-frequency statistics. The ONS could seek to incorporate web-scraped data into the official CPI and RPI but this decision will depend on gaining access to supermarket data of sufficient quality.[36]

Indonesia carries out a continuous household socioeconomic survey, the SUSENAS, meaning it can produce consumption poverty data every quarter (Alkire, 2014). In 2011, the core part of the survey was conducted quarterly – it sampled 75,000 households and was representative to the provincial level. Yearly data, with a cumulative sample of 300,000 households, are representative to district level (Natih, 2014).

Colombia's National Statistics Office (DANE), has embarked upon an ambitious three-year modernisation strategy in which big data play a key part.[37] The plan involves: developing pilots for monitoring inflation through web-scraping, and poverty through cell-phone activity, for instance; data-sharing partnerships with other public administrations and private corporations (such as the National Police and the bus transit system operator Transmilenio on crime prediction notably); and dedicated training programmes to strengthen internal staff capacities.

4.2 The political economy of data innovation

It is instructive to consider why countries that have made investments in data have done so; in other words, to look at the political economy of their choice. In Rwanda, the President has said that statistics are a priority, and this has tallied with the exigencies of donors' budget-support operations conditional on the tracking of development results (Krätke and Byiers, 2014).

In Mexico, the impetus came from joining the Open Government Partnership (OGP), also at the instigation of the president, who issued a presidential mandate stating that all government data would be public – the only exception being for security reasons. The country was a founding member of the OGP, and is now co-chair.[38]

In these countries, a data revolution is not an idea, or something to be agreed internationally – it is already happening. As Cosmas Ochieng, executive director of the African Centre for Technology Studies says: 'If you want to know about the demand side, just look at this. Governments are not finding what they need in official statistics. Traditional statistics might be good for research, but they can't or don't tell you the kind of information that governments need to know how to implement public policy' (see Box 12, overleaf).

However, it is interesting to note that in Rwanda, paradoxically, data are being opened, but space for civil society groups is being closed. Similarly, the Mexican president's personal commitment to transparency sits uncomfortably with his domestic reputation for overseeing a troubled regime (*The Economist*, 2014a). This underlines the fact that data alone are not sufficient to bring about change; a key factor is how these data are used.[39]

35 For information on the methodology for the reclassification, see Secretaría de Salud (2011).

36 See ten Bosch and Windmeijer (2014) for discussion of advantages and difficulties of webscrapers, also known as internet robots.

37 http://www.dane.gov.co.

38 http://www.opengovpartnership.org/country/mexico.

39 For further discussion, see Carother and Brechenmacher (2014).

Box 12: Kenya: how it became the Silicon Savannah

Kenya is charting the path for the data revolution in Africa. But how did it become the 'Silicon Savannah' as it is known by some? There have been three main factors behind its data leadership: a supportive government, the high rate of mobile phone penetration, and the growth of technology innovation spaces, although all roads seem to lead back to one man: Bitange Ndemo, Permanent Secretary in the Ministry of Information and Communication from 2005 to 2013.

Government policies

Over the last ten years, Kenya has been at the forefront of investing in telecommunications technology, open data, and encouraging the creation of technology innovation. With close ties to then President Kibaki, Ndemo was able to progress innovative government policies. He recognised that in order for Kenya to develop, it had to become more connected. Investment was focused on constructing cellular base stations, and now more than 80% of the Kenyan population has access to a mobile phone (World Bank, 2015a).

Ndemo encouraged the government into partnership with a company from the UAE to lay four fibre-optic submarine cables which resulted in Kenya becoming the African country with the 'highest bandwidth per person, the fastest speeds, and some of the lowest internet costs on the African continent' (Invested Development, 2014). This in turn lowered the barriers to entry for technology entrepreneurs and companies to develop mobile- and data-driven enterprises. In the face of opposition from many within government and the President's cabinet, Ndemo strongly advocated the creation of an open data policy in Kenya. Eventually, and with the support of President Kibaki, the Kenya Open Data Initiative was launched in 2011.

Mobile phone penetration and success of M-PESA

During the last ten years, the Kenyan government has also fostered a regulatory environment that has encouraged and promoted the development of technology-based innovations. Most notable of these is the mobile money platform, M-PESA. During its initial phase, the government and its financial regulation bodies took a hands-off but supportive approach (Robertson, 2015). Mobile money took off readily, laying the groundwork for mobile applications to be used in other sectors.

Technology innovation hubs

The explosive development of these data-driven mobile-based applications grew from innovative and entrepreneurial tech innovation centres, the most famous of which is the iHub, established in 2010. This co-working space has been credited for the creation of a number of highly successful initiatives driving the creation of a large segment of the country's unstructured data, including Ushahidi, Ma3Route and mFarm. The establishment of this tech community, paired with a supportive government, has helped to encourage leading technology companies such as IBM, Oracle, and Google to invest in Kenya. IBM chose to locate its Africa Research Lab in Nairobi, along with the $100 million Project Lucy, which will bring IBM's Watson cognitive computer system to Africa.*

** See IBM Research, IBM Brings Watson to Africa for Project Lucy (www.research.ibm.com/labs/africa/project-lucy.shtml).*

Source: Frosina, Wanda, and Mungwanya (2015).

4.3 A data revolution for citizens: reasons for optimism

Just as there is an ongoing data revolution for government, a data revolution for citizens is also underway. Here are just a few examples of data being leveraged for impact in a range of different ways and in different contexts.

In **Mumbai, India**, the state government wanted to expand the rail and road network, which meant resettling 18,000 families who lived along the railway line. Resettlement is generally very politically fraught, particularly when World Bank funding is involved, as was the case here. But the government asked a slum dweller's association to survey every dwelling. The association then used the information to negotiate with the Bank and the government. By 2008 all the families had been voluntarily resettled and, perhaps even more importantly, these baseline data were accepted as official data, giving the families security of tenure for the first time (D'Cruz and Mulayan, 2014).

Citizens' groups across Asia decided to set their own poverty line based on data from their own lives. Recognising the severe limitations of the $1.25-a-day definition of poverty, urban poor community groups, who had been elected by the local community **in Thailand, Nepal, Sri Lanka, the Philippines, Cambodia and Vietnam** assessed the minimum daily expenditure that would allow them to 'live decently and sufficiently and to maintain themselves as human beings'. Participants included the very poorest people, such as homeless people in Cambodia. In Thailand, the figure was found to be $4.74 (against the national poverty-line figure of $1.75) (Asian Coalition for Housing Rights, 2014).

In [certain] countries, a data revolution is not an idea, or something to be agreed internationally – it is already happening.

UNICEF, White Ribbon Alliance and World Vision, working together in the A Promise Renewed Partnership,[40] are collaborating with communities to produce scorecards of regionally disaggregated performance on a range of health indicators. Using a traffic-light system rather than numbers, the cards are being used to hold local authorities to account. In 2014, in **Cross River State, Nigeria**, the State Commissioner for Health was alerted that the indicator on the use of treated bednets by children was red, even though there was a very large stock of bednets in storage. After further investigation, it was discovered that 1.4 million treated bednets had not been distributed[41] due to lack of budgetary provision. The Commissioner pledged to find the necessary funds. Five months later, at the time of the next Partnership visit, three 18-wheeler trailer trucks were at the State Central Medical Store loading treated bednets for distribution. Of course, this success was not simply a result of the data, but how different groups came together to act upon them (Wild et al., 2015).

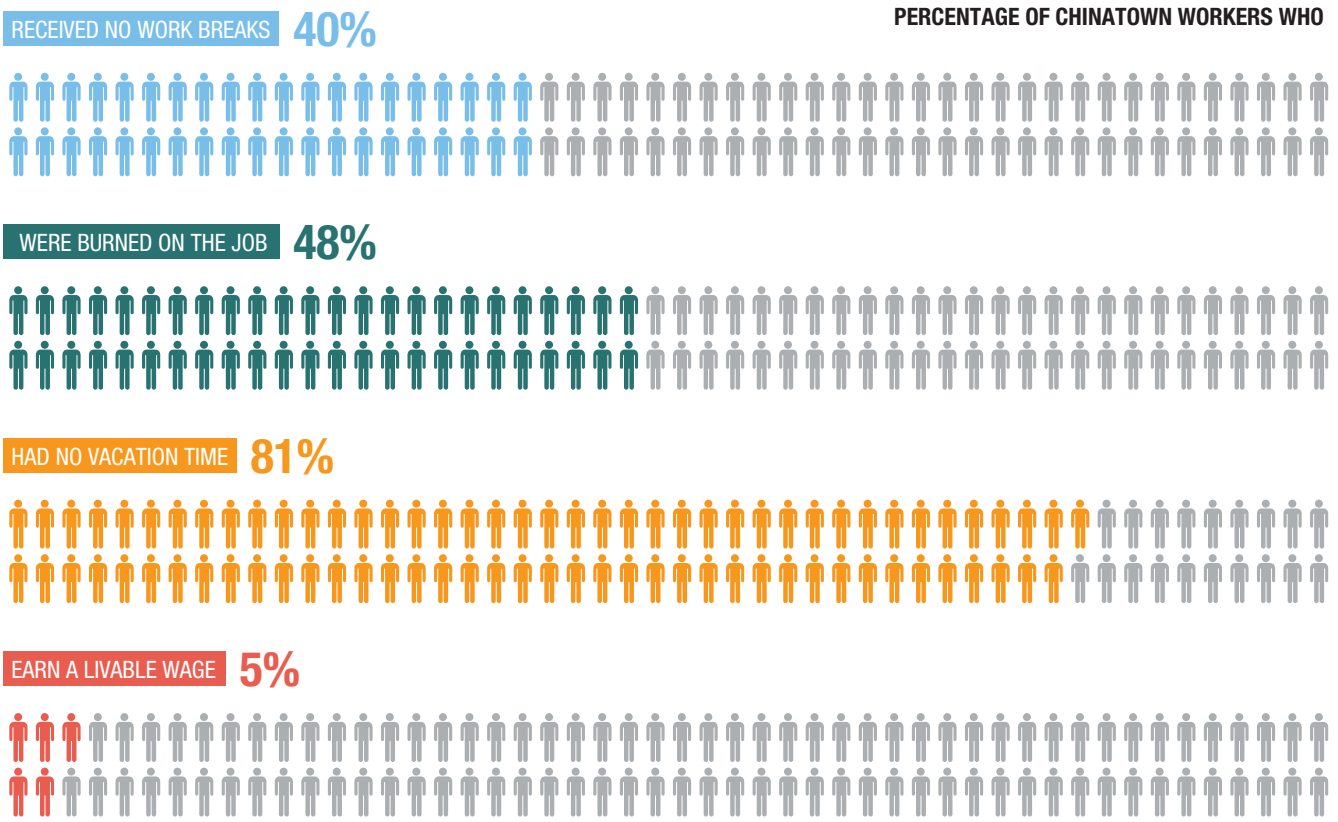
In **Uganda**, in 2012, a coalition of community organisations questioned the government on its failure to meet commitments to health spending, convincing it, in the end, to increase health allocation in the national budget. The coalition used its own district-level data, together with regionally disaggregated Ministry of Health data, to highlight severe shortfalls in staffing at clinics in different parts of the country, ascribing Uganda's poor performance in the MDGs to this under-investment. It ran a campaign which included sending SMSs to MPs and the media. In September 2012, legislators delayed the budget approval

Pre-test of rural household survey in Pakistan. Photo: © Pakistan Strategy Support Program.

40 A Promise Renewed is a global movement of public, private and civil society actors to end preventable child deaths by accelerating progress in maternal, newborn and child survival.

41 The bednets had been received from the Global Fund to Fight AIDS, Tuberculosis and Malaria.

Figure 5: Chinatown workers suffering abuses, San Francisco



Source: Adapted from Minkler (2014).

process by two weeks, until an additional US$19.8bn was allocated to the health sector for recruitment of health workers and enhancement of medical workers' salary packages (World Vision International, 2014).[42]

After the decline of the San Francisco garment industry in the early 2000s, in the **United States**, many of the city's Chinese immigrant labour force, a marginalised low-income community who spoke little English, had to compete for jobs in the local restaurant trade. These jobs were often defined by long hours, poor wages and workplace abuses (including verbal abuse, wage theft and denial of worker rights). The Chinese Progressive Association, San Francisco (CPASF), a local community group, asked the academic community to explore how they could capture data to empower workers and hold employers to account. Shaw San Liu of the CPASF says, 'We wanted to be able to use our resources to move forward a longer institutional plan to combat conditions and organise workers across the industry… and we knew we needed data to do that'.

The worker leaders conducted detailed interviews with 433 restaurant workers and observed 106 of the 108 restaurants in Chinatown. Results captured the endemic nature of labour violations and poor working conditions: half the workers reported being paid less than minimum wage; and half worked more than 60 hours a week (Figure 5). In response, in 2010, community leaders helped write a wage-theft ordinance, supported by the mayor. In 2013, the city sued Dick Lee Pastry Inc. for forcing employees to work 11-hour days, six days a week for less than $4 an hour. The restaurant settled with the city for $525,000 in back wages and penalties. In 2014 the survey data contributed to the winning of a $4m settlement with Yank Sing in 2014. The restaurant also agreed to further workplace changes, including wage increases, holiday and vacation pay, and paid healthcare for full-time employees.

42 Also based on author's interview with Hwa Yoo (Social Accountability Adviser) and Besinati Mpepo (Senior Policy Advisor Social Accountability), World Vision International.

Box 13: Freedom of Information requests flourishing across the world

Citizens around the world are making use of the evolution of 'right to information' laws (also known as 'freedom of information' or 'FOI') to gather data and information from the government that in the past might otherwise have been kept shielded from the public. This is helping to improve accountability of government behaviour and empowering citizens with greater knowledge and understanding of issues affecting them. At least 95 countries currently have these laws, meaning such citizen-government interaction can be found in both highly developed and fledgling democracies.

Figure 6: Identity of FOI requesters in Brazil and Finland

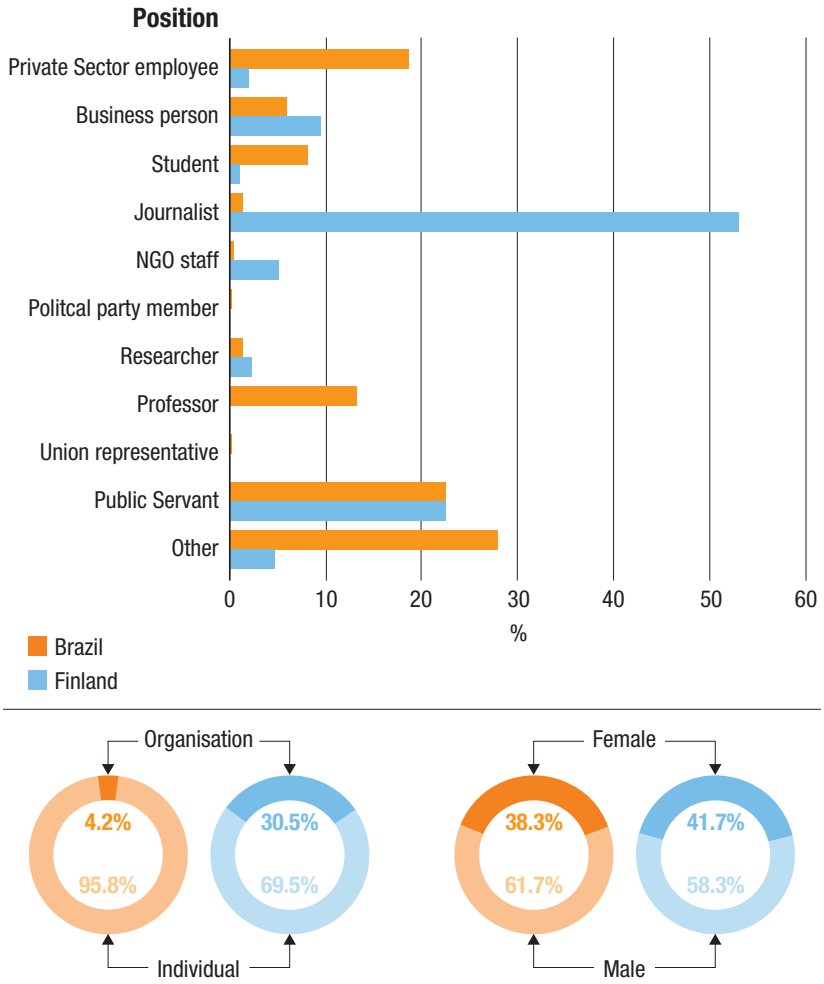


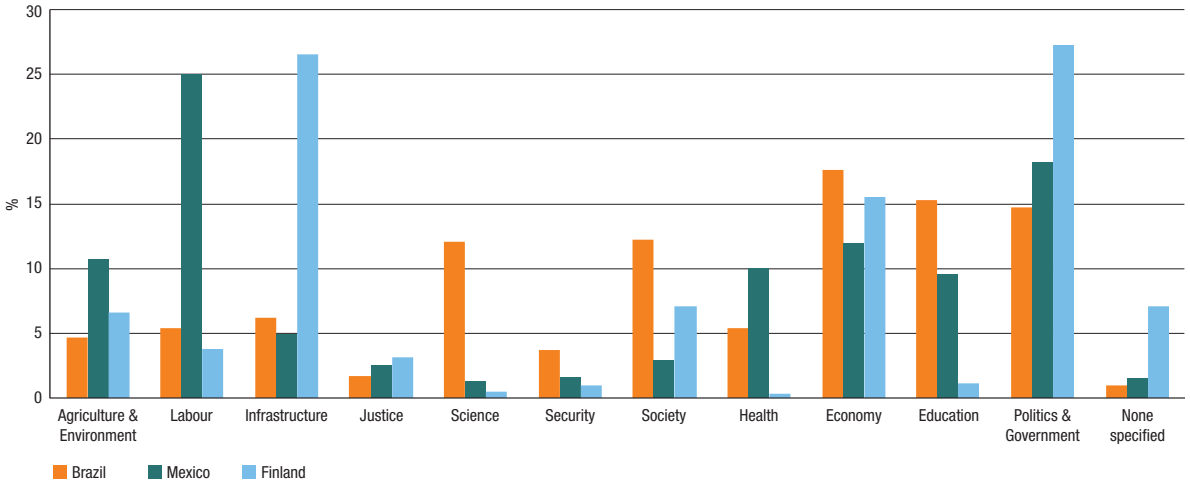
Figure 6 shows the background of FOI requesters in Brazil and Finland, two very different countries with very different civil societies. While in Finland over half of assessed requests came from journalists, only around 2% did so in Brazil. Students in Brazil appear to be much more active in making FOI requests than in Finland, as do university professors. Workers in Brazil also appear to be more active than businesses in requesting information from the government, whilst this relationship is reversed in Finland. The culture of FOI requests is slightly more gender neutral in Finland than in Brazil.

Interestingly, while over 99% of assessed requests in Brazil came from within the country, requests were still identified 28 other countries. In Mexico, discussed in Figure 7 (overleaf), FOI requests came in from a staggering 57 nations. Yet in our data from Finland, there wasn't a single identified request from outside the country.

The content of these requests, displayed in Figure 7, is also very different across countries. Labour issues make up a quarter of all FOI requests in Mexico but only 5% in Brazil and 4% in Finland. The economy and education are the top two concerns of FOI requests in Brazil, yet education receives very little attention in Finland. Given Finland's reputation for its outstanding education system (Pearson, 2013) versus the much ridiculed, lagging school-system in Brazil (*Financial Times*, 2015), this suggests that citizen's desire to inquire, using FOI requests, is positively associated with poor government performance.

Box 13: Freedom of Information requests flourishing across the world (continued)

Figure 7: Topics of FOI requests in three countries



Sources:

Authors' analysis of Brazil FOI requests (2014) from: http://www.acessoainformacao.gov.br/sistema/Relatorios/Anual/RelatorioAnualPedidos.aspx

Authors' analysis of Mexico FOI requests (2014) from: https://www.infomex.org.mx/gobiernofederal/home.action

Analysis of FOI requests from Finland (2012/2013) was produced by Open Knowledge Finland, assessing requests made via the Tietopyynto portal (http://tietopyynto.fi/). The authors do not claim that data displayed here is necessarily representative of all FOI requests made in each country, as some requests may not have been recorded via the sources above.

Box 14: Colombian rice growers and a democratic revolution

From 2007, yields in Colombia's rice-producing districts started falling dramatically: from 6 to 5 tons per hectare in five years. Rice is the staple food of almost a third of the population, particularly poorer communities, and the primary source of income of numerous small-scale subsistence farmers. A federation of rice growers (Fedearroz); the Centro Internacional de Agricultura Tropical (CIAT) and Colombia's ministry of agriculture together analysed crop-related data on an unprecedented scale.

The CIAT team used data-mining techniques to analyse information from annual rice surveys, harvest monitoring records, and agronomic experiments with changing the dates of rice sowing, as well as weather data from the National Institute of Hydrology. Results suggested that for several regions climate change was the problem, and that traditional farming practices needed to change too. The results are highly site-specific. In the town of Saldaña, for example, rice yields were limited mainly by solar radiation during the grain-ripening stage. In Espinal, the biggest problem was warm nights. This suggested that farmers in Saldaña could boost yields by aligning their sowing dates with sunnier seasons, whereas those in Espinal may have needed a variety more suited the local climate. 'When we use machine-learning techniques, we can explore non-linear functional relationships between various factors – temperature, radiation, rainfall and productivity', says CIAT agronomist Daniel Jiménez.

The analysis forecasted that the first growing season of the year would coincide with a severe dry period and that the window for planting crops would need to be delayed. A simple, site-specific message was articulated and communicated by Fedearroz to 170 farmers in Cordoba, a region in the north: 'don't plant in this sowing season'. To make sure the rice growers took the recommendation seriously, they gave pinpoint-accurate information to the farmers on either the ideal windows for planting or the best variety to grow.

'One farmer said to me that they acted on the research because it was based on their own data,' says Jiménez. This information helped these farmers avoid the drought and estimated economic losses of $3.6 million for that harvest. Now the CIAT team intends to scale up the initiative, using a similar approach with rice growers in Peru and Nicaragua through a partnership with the Latin American Fund for Irrigated Rice (FLAR).*

** Adapted from Cariboni, D. (2014); Russell (2014); Centro Internacional de Agricultura Tropical and UN Global Pulse (2015)*

5. A data revolution: the future

Survey agent at staff meeting in Bungoma, Kenya. Photo: © Lukas Bergstrom.

A fully realised data revolution – building on current innovations, and on the growing movements towards freedom of information, open data, and transparency – has the potential to go even further than the examples above have suggested. A future data revolution could completely change the way that citizens and governments interact.

First, it could contribute to the transformation of citizen-state relations. Data can be used to track service provision, enable citizens to reallocate local budgets, make changes in their communities, hold their governments to account and to participate better in democratic processes to ensure their needs and concerns count – often for the first time. A data revolution, in its most successful version, could mean marginalised people driving the agenda for their own progress.

Second, big data and new forms of data collection will give citizens new information they need to live better lives and earn more secure livelihoods. They can tell people the best time to avoid traffic, when best to plant crops, and which waterholes are free from arsenic, fluoride, iron, and parasites.[43] They will allow some people, for the first time, to register the birth of their baby who may then, as an adult, find formal employment and own land.

Third, data collected through new technologies can act as an early-warning system. Even if we do not know at the macro level the precise number of clinics or pharmacies that stock vital medicines, if people can alert their government via SMS to stock outs, this signals problems in a certain area, meaning that action can be taken before a full dataset is available. In the developed world, Fix My Street serves a

43 http://caddisfly.ternup.com/.

similar function.[44] In humanitarian situations, the real-time nature of data is potentially game-changing. As Robert Kirkpatrick from Global Pulse says: 'We've always had maps. Now we have people on them too.'

5.1 No automatic impact

With social accountability movements, there has been as assumption that generating information on budgets would arm citizens with the tools they need to lobby the government for change, and therefore that advocacy would happen, together with the ensuing change. However, a meta-analysis of 25 evaluations of field experiments has shown that this is not so easy. Jonathan Fox puts it elegantly: 'widely accepted, normatively appealing theories of change, summed up as 'sunshine is the best disinfectant' turn out to have uneven empirical foundations' (Fox, 2014).

Rakesh Rajani, former head of Twaweza – a citizen accountability organisation in east Africa, admits that many social accountability projects fizzle out, including some of Twaweza's, adding that for every one that goes right, ten are likely to fail. Rajani says: 'There's too little attention on old-fashioned human incentives, an automatic assumption that people will always give back feedback if you give them the tool. The old constraints of 'who am I to speak/it will be used against me/what's the point' are still there. These are problems of power and agency – they are the largest challenges for use of data-feedback. Just having new data or ways of analysing doesn't trump these constraints. If the government was non-responsive before, technology and data won't solve that problem or suddenly turn it into being more responsive. Data does not assure you that that voice will count.'[45]

It is clear that having timely data alone will not suffice to shift entrenched power imbalances. But, used as part of a dynamic, evolving feedback loop, data can empower and bring about change.

5.2 Towards a theory of change

Despite the above caveats, some of the insights on the role of information in social accountability movements can suggest some pathways for change for data (McGee, and Gaventa, 2010). Community groups will mostly not have access to governments, and an intermediary is generally needed to ensure that the advocacy messages, informed by data, reach the right audience and are taken seriously.

Jonathan Fox's analysis indicates that, unless a strategic approach is taken which coordinates coalitions and networks with reform processes already underway, access to information alone is unlikely to suffice (Fox, 2014). In addition, he points out that information needs to be perceived as actionable, which means that an enabling environment needs to reduce fear of reprisals (Fox, 2014).

A future data revolution could completely change the way that citizens and governments interact.

Finally, it is not likely that projects using data will get this right the first time. It will be a matter of testing, re-testing, adjusting and learning. Rajani says: 'The point here is not to experiment all day in boutique labs with little regard to impact, but rather to integrate experimentation and adaptation at the heart of how we implement at scale.'

5.2.1 Potential new technologies

Currently, the majority of big data methodologies entail using a mobile phone. The International Telecommunication Union estimates there are around 7 billion mobile subscriptions globally, although some people may have multiple subscriptions, while others may share phones. However, while mobile phone coverage is very extensive, it is not yet universal. The World Bank estimates that there are 1.4 billion people without a mobile phone subscription – although access rates to mobile phones tend to be significantly higher than subscription rates. A 2011 household survey in rural China estimated that 95% of people living in rural areas had access to a mobile, for example (World Bank, 2012).

The people most likely not to have access are the poorer, least-educated and female, or those living in the most remote areas.[46] However, coverage is increasing all the time, and efforts such as the White Spaces project in South Africa and Kenya are experimenting with access in the most geographically difficult places (Wanjiku, 2013).

Robert Kirkpatrick from Global Pulse says: 'We may not have 100% mobile phone coverage in Afghanistan for instance, but let's figure what we can do with the data to be of most use to the 80% that do have it, so that way we won't waste time once we do have universal coverage.'

But other sources of big data are emerging that could have the potential to include those who are otherwise off-grid. These include:

- **Geo-sensing.** This is already used extensively (see Box 16, page 41 on satellite data).
- **Community radio.** Even the poorest families are likely to have a radio. Pulse Lab Kampala is piloting

44 https://www.fixmystreet.com.

45 See Green, D. (2014) and Hagen (2013).

46 In Rwanda, for instance, phone owners were found to be 'considerably wealthier, better educated, and predominantly male', with mobile phones 'disproportionately owned and used by the privileged strata of Rwandan society' (Blumenstock and Eagle, 2012).

a programme which uses audio-to-text software to transcribe issues that people are discussing, and textual analytics to capture key words, and therefore emerging concerns, in a similar way to big-data analytics of Twitter feeds. The lab is also working to develop new speech models in the vernacular Ugandan languages of Luganda and Acholi.[47]

- **Postal data**. Even where individuals don't leave a data footprint, their interactions might do so. One billion letters are sent daily from 650,000 post offices, and each one generates around 20 data points. Global Pulse and the Universal Postal Union are examining these data to see what they can tell us about economic trends, migration, poverty and resilience.[48]

Box 15: The potential of data to transform India's Public Distribution System

India's Public Distribution System (PDS) is the world's biggest food security network. The system aims to deliver highly subsidised staple food grains such as rice and wheat, on a monthly basis, to two-thirds of India's 1.2 billion population – a staggering 800,000,000 people – via what are known as Fair Price Shops. The programme is believed by many politicians to be a key benefit to the country's most vulnerable groups, providing a minimum support price for a significant proportion of India's estimated 119 million famers.

The PDS is, however, highly controversial. Common criticisms include high levels of leakage, corruption, misallocation and mis-identification of beneficiaries. According to one assessment,* 67% of wheat and 41% of total food grains destined for distribution are siphoned off to the open market. Too often the wrong (that is, not the poorest) people are targeted – just 13% of households below the poverty line (BPL) access their full entitlement, and a third don't get anything from the Fair Price Shops. It is also exceptionally expensive to administer: the Ministry of Consumer Affairs, Food & Public Distribution estimates running costs to be $20bn.**

Technology and data management are now being used to address these problems. In Chhattisgarh, where reforms started in 2000, a central beneficiary database has been set up and beneficiaries now use a bar-coded ration card to collect their goods. Monthly allocations are automated using a web-based application. Delivery orders, receipts and movement of commodities between distribution centres are carried out through application software. SMS alerts are sent to registered beneficiaries for a designated Fair Price Shop whenever commodities are dispatched from central stores. Citizen vigilance committees have been established and a grievance-redressal system, via a toll-free number, allows users to complain as needed (Puri, 2012).

Drèze and Khera (2013) note a correlation between a well-functioning PDS and reductions in both poverty-headcount ratios and calorie intake, and a reduction in skipped meals. A comparison of a well-performing (Chhattisgarh) and poor-performing (Bihar) PDS, demonstrates the impact that reform can have.***

Table 2: The PDS in Bihar and in Chhattisgarh

	Bihar	Chhattisgarh
Proportion of BPL households who did not get any food grains from the PDS in the last 3 months (%)	35	0
Average food grain purchase of BPL households from the PDS in the last 3 months. As a proportion of entitlements (%)	45	95
Proportion of BPL respondents who said that they 'normally' get their full PDS entitlements (%)	18	97

Source: Drèze and Khera (2013).

** See Khera (2011).*

*** Press Information Bureau, Government of India (2013). The estimated cost of US$20 billion also includes the cost of other welfare programmes associated with the NFSA.*

**** For further analysis of the benefits of the PDS reform in Chhattisgarh, see Krishnamurthy, et al. (2014).*

47 http://pulselabkampala.ug/radiomining/.

48 See Miguel Luengo-Oroz, speaking at the TedXNations event in Geneva on 11 December 2014 (http://www.unglobalpulse.org/TedX-UPU-big-postal-data).

- **Drone data.** Drones are being used successfully to track the movement of endangered tigers in the Panna Tiger Reserve in Madya Pradesh in India, and the government is now considering the use of drones in 10 other areas rich in biodiversity (Ramadurai, 2014). Drones could also be used to gather information on the very hardest-to-reach populations, such as: nomadic pastoralists; in disaster relief efforts (Tode, 2014); or to map land rights.[49]

But more potentially transformative than the collating of data that matters to ordinary people is the capacity of people to gather data on themselves, or on issues relevant to their own lives. The ultimate in demand-driven development data is the poor or vulnerable gathering – and using – those data themselves.[50]

The Community-Based Monitoring System (CBMS) project, started in the Philippines and now implemented in 22 countries[51] across Africa, Asia and Latin America, entails community members and local officials tracking poverty and development issues at household level. This information has proved crucial in many places for local budgeting interventions addressing specific group needs. For example, it is extremely adept at supporting highly accurate targeting for social programmes for the poorest. In the Philippines, street-by-street identification of the most deserving recipients for conditional cash transfers (CCTs) has led to particularly good results in improving school attendance and immunisation, the conditions of these CCTs (Reyes, 2014). In fact, nearly 50% of all 'bangarays' (villages) in the Philippines use CBMS (World Vision, 2014). This is a cheap way of gathering data: it costs around US$0.70 per household in the Philippines, and only US$0.30 in Vietnam (World Vision, 2014).

Box 16: Satellite data

Strikingly, one source of big data that even the most stretched national statistics office in developing countries report using is satellite images, although even 40 years ago this would have seemed an impossibly sophisticated technological frontier. The use of satellites to predict weather is well known. But satellite data can also map differences in the Earth's surface so precisely that we can calculate how much water is stored in even the most isolated of aquifers (Sullivan, 2015), or analyse the quality of soil (Lenhardt, 2015).

Light emissions picked up by satellites are also being used to proxy poverty levels and track GDP growth to supplement national accounting in data-poor countries (Henderson et al., 2012). This finding has been validated elsewhere: for example, Chen and Nordhaus (2010) and Olivia et al (2014) use 'gold standard' data on electrification and economic growth for 5,000 sub-districts in Indonesia between 1992 and 2008. But there is also evidence that this relationship can fade once the penetration of electric lighting approaches saturation.

49 The United Arab Emirates, Drones for Good Award, see (https://www.dronesforgood.ae/en/award/finalists/project-amer-autonomous-mapping-and-evaluation-robot). For a discussion of privacy issues surrounding the use of drones, see *The Economist* (2015).

50 Although the demand for data collection is not always clear (often this may be at the instigation of an NGO or a donor).

51 See Partnership for Economic Policy, Country Project Profiles (www.pep-net.org/country-project-profiles).

6. Getting the innovation right

Big data measure what people actually do, rather than what they say they do. As such they can in some ways be more robust and subject to less bias than traditional forms of data. However, big data are also for the most part untried, and so investing in them should not be done unthinkingly.

6.1 Interpretation

Big data will only add value by increasing our understanding of the world if used with an understanding of context. In a developing country, big data can measure in real time and with highly localised detail. Because of this it can, for instance, give insights into the relationships between different communities or on how people in a certain area respond to sudden disasters like flooding, as in Haiti and Indonesia.[52] However, data removed from their original context and analysed without an understanding of that context could give rise to inaccurate interpretations. This could be more unhelpful than not having data at all, as it may lead to governments (or others) making the wrong decisions based on what they understand to be the evidence (Box 17).

Social media for example, will give insights only into the conversations of a very small – and unrepresentative – sample of the population even in developed countries, let alone in developing ones. However, this can be mitigated by focusing research in a part of the world where lots of people use that form of social media. A Global Pulse case study tracking food price inflation using Twitter data did so in Jakarta, the city with the largest Twitter presence in the world, for example (Box 18, overleaf).

Another way in which data scientists deal with this issue of selection bias is to correct for it, that is to make assumptions about the population that is excluded, based on household survey data, and then adjust their figures accordingly.[53] But although big data can measure new aspects of life, in other ways it is less rich than traditional data. It doesn't tell us about how these aspects interact with one another (for example, how the dimensions of poverty may be linked and interact with one another at a household level), which is vital for targeting poor households and for delivering services effectively.

6.2 Integrating different kinds of data

Although this paper started by highlighting inadequate data, paradoxically in some cases accuracy is not what is needed. Users, including civil society groups and governments, may value speed and/or granularity over precision, depending on the use and the circumstances. When we use Trip Advisor, we do not expect formal peer reviews of restaurants or hotels for instance. This may be a

Box 17: Big data and spurious correlations: one way to get interpretation wrong

Big data gives us a whole new set of datasets for hypothesis testing. Used correctly, this process can drive significant innovation.

But if you have a big enough dataset, you can find correlations of almost anything. A website called 'spurious correlations' run by a Harvard law student shows this to brilliant effect. He lists just under 40,000 correlations you can make by putting together data from the US Census and the Centers for Disease Control and Prevention (CDC) Wonder databases (also US data). You can, for instance, correlate the per capita consumption of cheese in the US with the number of people who die by becoming tangled in their bed sheets (correlation of 0.947091), or US crude oil imports from Norway with drivers killed in collision with a railway train (correlation of 0.954509).

The example is clear – correlation does not equal causality. Research is needed to develop new methods and algorithms that can handle big data and that address issues of how you compare datasets and draw robust inferences. Unless this happens, it is possible that we find correlations that are at best artefacts and are at worse harmful – in that they could be used to (mis-)inform policy.

Source: tylervigen.com.

52 http://haitidata.org/maps/133 and www.futuregov.asia/articles/6057-open-dataset-of-the-week-flood-prone-areas-in-jakarta.

53 Using a so-called difference-in-differences approach. See Guerrini (2014).

Box 18: Mining Indonesian Tweets to understand food price crises

Since 2010, Indonesia has witnessed substantial increases in food prices: the price of rice increased 51% between December 2009 and February 2012. With more than 20 million Twitter user accounts in Jakarta, a wealth of data is being produced daily. UN Pulse Lab Jakarta sought to explore this source to monitor conversations regarding food security. The research shows that the automated monitoring of public sentiment on social media, combined with contextual knowledge, has the potential to be a valuable real-time proxy for food-related economic indicators (UN Global Pulse, 2014).

Pulse Lab Jakarta analysed Twitter conversations discussing food price increases between March 2011 and April 2013. Taxonomies, that is groups of words and phrases with related meanings, were developed in the Bahasa Indonesia language to identify relevant content. A classification algorithm was trained to categorise the extracted tweets as positive, negative, confused, or neutral to analyse their sentiment. Using simple time-series analysis, the researchers quantified the correlation between the volume of food-related Twitter conversations and official food inflation statistics.

A relationship was found between retrospective official food inflation statistics and the number of tweets speaking about food price increases. Moreover, upon analysing fuel price tweets, it was found that perceptions of food and fuel prices were related. In particular, the researchers found a significant correlation (r=0.58) between the two topics, suggesting that even potential fuel price rises affect people's perception of food prices.

Source: UN Global Pulse (2014).

source of tension between statisticians, whose incentives are to produce high-quality statistics, and big data providers who can produce data quickly but not always cleanly.

But these two demands may be complementary rather than mutually exclusive. For instance, there's currently a lag of around two years before official aid data are published by the OECD Development Assistance Committee (DAC). That the DAC has ensured a consistent definition of Official Development Assistance, and that the official numbers can be tracked over time is a very valuable gold standard, particularly to voters who want to hold their donor governments to account. However, numbers that are more than a year out of date are not helpful for developing-country citizens' groups wanting to track budgets. The International Aid Transparency Initiative (IATI) on the other hand provides forward-looking management-information data as donors release them. Around a third of the donors and international NGOs report at least quarterly.[54] The trade-off is that not all donors currently report to it. In other words, new data serve a different function from traditional data forms, but both are still needed.

Combining traditional forms of data, and new forms such as big data, is known as interoperability (Box 19). New methodologies will need to be devised to do this (Box 20) and these will vary across sectors. Google chief economist Hal Varian has written extensively on new tools and techniques for using big data in econometric analysis, for example (Varian, 2014).

6.3 Reaching the marginalised

The data revolution can be deemed successful only if it has improved the lives of all, not just of a technologically enabled elite. Given the failure of many seemingly radical new technologies to change structural power dynamics in the past (Graham, 2014), the fears of civil society groups that a data revolution might widen and entrench a digital divide are perhaps well founded.

When members of the New Zealand Data Futures Forum first talked to Maori leaders, they found clear concerns. 'Many Maori do not perceive themselves as having benefitted much from the collection and use of data. They perceive a real and immediate risk of greater data availability being used for ethnic profiling to their detriment. Despite widespread demands on them for data in the past, the data seems to be rarely used in ways that might benefit them… Collection, storage and use of data often occur in ways that do not respect Maori *tikanga*.'[55]

But when new forms of data are combined with better use of traditional surveying techniques, rather than

Whatever the political context, as new forms of data emerge, so too do vitally important issues around the concentration of knowledge and power.

54 As of February 2015 (http://dashboard.iatistandard.org/timeliness.html)

55 Guidelines for daily life and behaviour in Maori culture. Authors conversation with James Mansell, Director of Innovation, Ministry of Social Development NZDFF.

pushing the currently marginalised further off the grid, they have the potential to enable, for the first time, the counting – and therefore the inclusion – of disadvantaged communities and populations. This involves considerable challenges. In some contexts, increasing surveillance of the already-marginalised or discriminated-against is potentially very problematic. Hypothetically, ethnic-minority communities could be targeted by police using predictive technologies that use data analytics to anticipate where and by whom crimes are going to be committed.[56] This issue has been raised in the context of the Obama Administration review of big data and privacy (The White House, 2014). Safeguards in the digital world will be very important to protect those already vulnerable in the real world, as we explore below.

6.4 Ownership, power, privacy and citizens' rights

Some have dismissed fears of the data revolution as a threat to privacy and a catalyst to worsening other power imbalances as a luxury that only advanced countries can afford. Alarm about invasive CCTV cameras, the sharing of National Health Service data in the UK, the IS National Security Agency spying in Germany, overzealous marketeers, and Facebook's social experiments, are – to borrow a trope from social media – #firstworldproblems.

In many poor countries, the problem lies not in the government and private sector knowing too much about people, but in the fact that so little is known. The World Economic Forum coins the term 'underknown' (WEF, 2015). But however scarce timely, usable data is today in developing countries, the private sector is forging ahead with new applications of technology that mean that increasingly complex power dynamics are inevitable in all countries North or South, and will need to be resolved.

Box 19: Combining new and old forms of data

Interoperable data can be used together in new ways not foreseen when they were collected. This can be achieved in two ways. Agreed-upon standards for the fundamental organisation of the data – semantics, structure and coding – can be applied to ensure that the data themselves are compatible and can be pooled and analysed together. The alternative is to move the compatibility requirement to a higher level and develop statistical methods that accept either data in various forms or already-prepared summaries of data, e.g. indicator values with uncertainty intervals, and then produce an overall summary or indicator value based on those diverse inputs. The second approach is more flexible and pragmatic but more ad hoc, requiring a new set of methods for each situation.

Note: Box contributed by Sam Clark, University of Washington.

Indeed, arguably, ownership and privacy questions are even more important in developing countries that are non-democratic or have otherwise repressive regimes, because the penalty for getting it wrong may be more serious than in advanced countries, with ethnic cleansing at the extreme range of the spectrum of negative potential outcomes. Whatever the political context, as new forms of data emerge, so too do vitally important issues around the concentration of knowledge and power. We may be uncomfortable with Google knowing so much about us, but what happens when the government has even more information? In particular, what happens to individual freedom in authoritarian regimes or when bureaucracies go awry? It is far from inconceivable that poor privacy practices could mean highly sensitive data about people's ethnicity could be used for nefarious purposes (see Box 20, overleaf).

56 See Data and Society Research Institute (2014).

Box 20: Some possible solutions

Solutions to privacy and data ownership questions lie largely outside the scope of this paper.* However, there are some potential partial solutions to the questions of power imbalance. One that appears to have the most potential to allow citizens to realise the benefits of data, while at the same time protecting their rights, is a **data commons**, whereby a government or private sector monopoly on data is broken by individuals giving permission for their data to be shared with whomever they choose, to their benefit. Tools for managing a personal data bank already exist, such as the OPEN PDS platform.**

But there are two key problems with this. First, people may not understand what they are signing, and second, it is not possible to guarantee today that technological advances will not render any current privacy assurances obsolete in future.*** Therefore, an appropriate regulatory environment will be necessary. Alex Pentland of MIT has proposed a **new deal on data**, whereby there would be workable guarantees that the data needed for public goods are readily available, while at the same time protecting citizens. The principles of this new deal have now been enshrined in legislation (in the US in the Consumer Bill of Rights, and in Europe in the EU Declaration of Data Rights) (*Edge*, 2012).

In East Africa, the African Centre for Technology Studies is working to influence the African Union to introduce a model law with guidelines and principles on privacy akin to the above, in the expectation that countries would take this into their own national legislature. The ultimate way to ensure privacy may be an old-fashioned one: ensuring that there are a range of different people with differing incentives around the regulatory table, and erring on the side of caution.

** See for instance WEF (2015) for a more detailed discussion.*

*** http://openpds.media.mit.edu/#philosophy.*

**** See the work by Yves-Alexandre de Montjoye and Alex Pentland (MIT) on anonymisation. They have shown that anonymous credit-card-transactions data can be reverse-engineered to identify individual's transactions (Berinato, 2015).*

Conclusion

Substantial data gaps across a range of sectors are hindering governments and citizens in solving problems of poverty, inequality and environmental sustainability. Because internationally comparable household surveys are conducted only intermittently, we have only cross-national snapshots of the world taken at three- or five-year intervals. And because data sources leave out – by design and in practice – the world's most vulnerable people, and frequently the issues of importance to them, neither governments not citizens have access to the information they need to support progressive change.

But progress is underway. Gaps are being filled by improving traditional statistical sources, such as censuses, household surveys and birth and death registries. Some countries are now conducting continuous surveys, providing a constant stream of information that reflects the dynamic nature of development. Gaps are also being filled by big data such as satellite images. And in the future, new sources of data gathered from interactions via mobile phones and the internet will reveal more about a wider range of people – with the potential to allow, for the first time, everyone in society to stand up and be counted, literally and metaphorically.

However, data will not necessarily lead to better policy-making and more accountability. Learning will be iterative: a range of programmes, projects and pilots experimenting with new uses and forms of data will be needed to find what works and to adapt to the specific context in a country or province. This approach will also mean that processes and ideas can be re-shaped as new technologies emerge.

This is entirely appropriate. The data revolution, while discussed in global fora, will actually happen at the national, regional and hyper-local levels – that is, as close to the needs and users as possible. Country-level data compacts, currently under discussion (Glassman, 2014), may be a good way to foster this, by bringing together governments, users such as citizens' groups and academics, donors and the private sector. Together, these stakeholders can work out the priority gaps for each country, and develop a road map to finance and deliver the change needed.

What is certain, however, is that the data revolution will only gather pace. Governments and citizens alike will need to harness its power to ensure that its social benefits are felt by all.

Ismail on mobile telephone in Ta Kuti village, Nigeria. Photo: © Arne Hoel / World Bank.

References

Abraham, S. (2014) Availability and accessibility of government information in the public domain. The Centre for Internet and Society.

Alkire, S. (2007) 'The missing dimensions of poverty data' (Guest Editor: Emma Samman), Oxford Development Studies 35(4): 347-359.

Alkire, S. (2014) Towards frequent and accurate poverty data. Sustainable Development Solutions Network.

Alkire, S. and Samman, E. (2014) Mobilising the household data required to progress towards the SDGs. Oxford: Oxford Poverty and Human Development Initiative.

Asian Coalition for Housing Rights (2014) *Newsletter of the Asian Coalition for Housing Rights* 19, September 2014 (www.achr.net/upload/downloads/file_16102014142111.pdf).

Bellagio Big Data Workshop Participants (2014) 'Big data and positive social change in the developing world: a white paper for practitioners and researchers'. Oxford: Oxford Internet Institute.

Berinato, S. (2015) 'There's no such thing as anonymous data'. Harvard Business Review.

Berkowitz, E. and Paradise, R. (2011) 'Innovation in government: Kenya and Georgia', McKinsey Quarterly (www.mckinsey.com/insights/public_sector/innovation_in_government_kenya_and_georgia).

Bhargava, R. et al. (forthcoming). 'Data literacy in the age of big data: imperatives, impediments, and requirements'. Data Pop Alliance.

Bird, K. (2007) 'The intergenerational transmission of poverty: an overview.' London: Overseas Development Institute.

Blas, J. (2014) 'Uganda joins African wave of GDP rebasing and gets 13 per cent boost', *Financial Times,* 1 December.

Blumenstock, J. and Eagle, N. (2012) 'Divided we call: disparities in access and use of mobile phones in Rwanda', Information Technology and International Development 8(2): 1-16.

Bosch et al. (2012) 'Taking stock of nine years of implementation of Seguro Popular in Mexico: lessons for developing countries'. IDB.

ten Bosch, O. and Windmeijer, D. (2014) 'On the role of internet robots for official statistics'. Paper Prepared for the Meeting of the Management of Statistical Information Systems, Dublin, Ireland and Manila, Philippines, April 2014.

Buettner, T. and Gerland, P. (2008) 'Preparing population estimates for all countries of the world: experiences and challenges', New York: United Nation Population Division.

Buvinic, M., Furst-Nichols, R. and Koolwal, G. (2014) 'Mapping gender data gaps', Data2X.

Cariboni, D. (2014) 'Colombia rice growers saved from ruin after being told not to plant their crop', *The Guardian,* 30 September.

Carother, T. and Brechenmacher, S. (2014) *Closing space: democracy and human rights support under fire*. Carnegie Endowment for International Peace.

Carr-Hill, R. (2013) 'Missing millions and measuring development progress', World Development 46: 30-44.

Centro Internacional de Agricultura Tropical and UN Global Pulse (2015) *Data for Climate Action Challenge* (www.unglobalpulse.org/data-for-climate-action-2015).

Chandy, L. (2013) 'Counting the poor: methods, problems and solutions behind the $1.25 a day global poverty estimates'. Development Initiatives and Brookings.

Chandy, L. and Kharas, H. (2014) 'The poverty of poverty data'. Brookings.

Chen, S. and Ravallion, M. (2010) 'The developing world is poorer than we thought, but no less successful in the fight against poverty', Quarterly Journal of Economics 125(4): 1577-1625.

Chen, X. and Nordhaus, W. (2010) 'Using luminosity data as a proxy for economic statistics', Proceedings of the National Academy of Sciences of the United States of America 108(21): 8589-8594.

Coalition for Evidence-Based Policy (2012) 'Rigorous program evaluations on a budget: how low-cost randomized controlled trials are possible in many areas of social policy'. Coalition for Evidence-Based Policy

Croke K., Dabalen, A., Demombybes, G., Giugale, M. and Hoogeveen, J. (2012) 'Collecting High Frequency Panel Data in Africa Using Mobile Phone Interviews'. Washington DC: World Bank.

Data and Society Research Institute (2014) *Event Summary: The Social, Cultural, & Ethical Dimensions of 'Big Data'*. New York: Data and Society Research Institute.

D'Cruz, C. and Mulayan, G. (2014) 'Community-led baseline surveys as a tool for the resettlement and rehabilitation (r&r) component for the Mumbai Urban Transport Project 2 (MUTP2)'. Slum Dwellers International.

de Benítez, S. (2007) 'State of the world's street children: violence'. London: Consortium for Street Children (UK).

Demombynes, G. and Sandefur, J. (2014) 'Costing a data revolution'. Washington DC: Center for Global Development.

Department of Business, Innovation and Skills (UK) and Ordnance Survey (2014) 'Assessing the value of OS Open Data to the economy of Great Britain' (www.nationalarchives.gov.uk/documents/meetings/20140128-appsi-what-is-the-value-of-open-data.pdf).

Development Initiatives (2014) 'Using resource tracking and feedback to enhance accountability and resource effectiveness'. Development Initiatives.

Dodgeon, S. (2013) 'Every mother counts: reporting health data by ethnicity'. London: Health Poverty Action.
Drèze, J. and Khera, R. (2013) 'Rural poverty and the public distribution system', Economic and Political Weekly XLVIII (45 & 46): 55-60.
Dykstra, S., Kenny, C. and Sandefur, J. (2014) *Global absolute poverty fell by almost half on Tuesday*. Centre for Global Development.
The Economist (2010) 'The Indian census and caste: caste in doubt', 10 June.
The Economist (2012a) 'The Food and Agriculture Organisation, the Finest Traditions of Australian Diplomacy', 3 October.
The Economist (2012b) 'Hunger, not a billion after all', 10 October.
The Economist (2014a) 'Protests in Mexico: ¡Justicia!', 6 December.
The Economist (2014b) 'A census in Myanmar: too much information' 22 March.
The Economist (2015) 'Drones and privacy: a looming threat', 19 March.
Edge (2012) 'A conversation with Alex Pentland: reinventing society in the wake of big data'. 30 August.
Espey, J. et al. (2015) 'Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development' (http://unsdsn.org/needsassessment).
Fox, J. (2014) 'Social accountability: what does the evidence really say?' Washington DC: The Global Partnership for Social Accountability.
Frosina, N., Wanda, S. and Mungwanya, R. (2015) 'Kenya Case Study', unpublished background paper prepared by the Africa Centre for Technology Studies.
Fung, A., Russon Gilman, H. and Shkabatur, J. (2010) *Impact case studies from middle income and developing countries: new technologies*. London: Transparency and Accountability Initiative.
Giugale, M. (2012) 'Fix Africa's statistics', Huffington Post.
Glassman, A. and Ottenhoff, J. (2014) 'Better data in Rwanda'. Center for Global Development.
Glassman, A. (2014) 'Data revolution from the bottom up'. Center for Global Development.
Government of Pakistan (2014) 'Pakistan Economic Survey 2013-14'. Chapter 15, Poverty and social safety nets.
Graham, M. (2014) 'Inequitable distributions in internet geographies: the global South is gaining access but lags in local content.', Innovations 9(3-4); 17-34.
Granoff, I. et al. (2014) 'Targeting zero zero: achieving zero extreme poverty on the path to zero net emissions.' London: Overseas Development Institute.
Green, D. (2014) 'What are the limits of transparency and technology? From three gurus of the openness movement (Eigen, Rajani, McGee)' From Poverty to Power [blog], 7 April (http://oxfamblogs.org/fp2p/what-are-the-limits-of-transparency-and-technology-the-thoughts-of-three-gurus-of-the-openness-movement).
Guerrini, F. (2014) *Using big data to understand migrations*. Forbes.
Hagen, E. (2013) 'The funding conundrum: why funding for ICT development needs to change'. GroundTruth Initiative [blog] 9 October (http://groundtruth.in/2013/10/09/the-funding-conundrum-why-funding-for-ict-for-development-needs-to-change).
HAI (Help Age International) (2014) Global Age Watch Index 2014: Insight Report. London: Help Age International.
Head, B. (2008) 'Three Lenses of Evidence-Based Policy Making', *Australian Journal of Public Administration* 67(1): 1-11.
Health Metrics Network (n.d.) 'Technical note on the costs of alternative approaches to collecting population and vital event dates'. World Health Organisation.
Henderson, J.V., Storygard, A. and D.N. Weil (2012) 'Measuring economic growth from outer space', American Economic Review 102(2): 994-1028.
Henry, J. (2012) 'The price of offshore revisited: new estimates for "missing" global private wealth, income, inequality, and lost taxes'. Tax Justice Network.
IBP (International Budget Partnership), DFI (Development Finance International), Oxfam America (2014) From Numbers to Nurses: Why Budget Transparency, Expenditure Monitoring and Accountability Are Vital to the Post-2015 Framework.
IEAGDRSD (Independent Expert Advisory Group on a Data Revolution for Sustainable Development) (2014) A world that counts: mobilising the data revolution for sustainable development. New York: United Nations.
IMF (International Monetary Fund) (2014) Kenya: 2014 Article IV Consultation – staff report; press release; and statement by the executive director for Kenya. Washington DC: IMF.
Instituto Nacional Des Estatisticas (2009) Final Report of the Multiple Indicator Cluster Survey, 2008. Mozambique.
Invested Development (2014) Enabling internet connectivity in Africa. Invested Development.
Järv, O. et al. (2012) Mobile phones in a traffic flow: a geographical perspective to evening rush hour traffic analysis using call detail records. PLOS One.
Jackson, J. (2014) Census data can empower all communities. Mizzima.
Jooste, I. (2013) 'Audience reception of news infographics: cultural and generational considerations'. Paper for Inter-Disciplinary.Net 7th Global Conference 2013.
Khan, A. et al. (2015) Progress under scrutiny: Pakistan's route to poverty reduction. London: Overseas Development Institute.
King, G. et al. (2009). 'Public policy for the poor? A randomised assessment of the Mexican universal health insurance programme', Lancet 373(9673): 1447-1454.
King, G. (2014) 'Why "Big data" is a big deal', Harvard Magazine.
Klasen, S. (2004), 'Gender related indicators of well-being', Discussion Paper 102, Goerg-August Universitat, Ibero-Amerika Institut fur Wirtschaftsforschung, Goettingen.

Klein, R. and Maybin, J. (2012) Thinking about rationing. London: The King's Fund.
Krätke, F. and Byiers, B. (2014) The political economy of official statistics implications for the data revolution in sub-Saharan Africa. Maastricht: ECDPM.
Kresge, K. (2007) 'HIV prevalence estimates: fact or fiction', IAVI Report 11(4).
Krishnamurthy, P., Pathania, V. and Tandon, S. (2014) 'Public distribution system reforms and consumption in Chhattisgarh: a comparative empirical analysis', *Economic and Political Weekly* 49(8): 74-81.
The Lancet Correspondence (2015) 'GBD 2013 and HIV incidence in high income countries', Lancet (www.thelancet.com/journals/lancet/article/PIIS0140-6736%2814%2960844-8/abstract).
La Porta, R. and Shleifer, A. (2014) 'Informality and Development', Journal of Economic Perspectives 28(3): 109-126.
Lenhardt, A. (forthcoming 2015) A measurable difference: Lessons from measuring development progress in 26 country case studies. London: Overseas Development Institute.
Letouzé, E. and Jutting, J. (2014) Official statistics, big data and human development: towards a new conceptual and operational approach. Data Pop Alliance and PARIS21.
Lopez, A. et al. (2007) 'Keeping count: births, deaths, and causes of death' (Comment), The Lancet online 29 October 2007.
Lucci, P. (2014) An urban dimension in a new set of development goals. London: Overseas Development Institute.
Mahtani, S. and Myo, M. (2014) 'U.N. Group Criticizes Myanmar Over Census Government Doesn't Collect Data on 'Rohingya' Muslims, Underscoring Ethnic Tensions'. Wall Street Journal.
Marcoux, A. (1998) 'The feminisation of poverty: claims, facts and data needs', Population and Development Review 24(1): 131-139;
Matheus, R. and Manuella, M. (2014) Case study: open government data in Rio de Janeiro City. Open Research Network
McGee, R. and Gaventa, J. (2010) Synthesis Report Review of Impact and Effectiveness of Transparency and Accountability Initiatives. Sussex: Institute of Development Studies.
McKinsey Global Institute (2013) Open data: unlocking innovation and performance with liquid information. McKinsey Global Institute.
Melamed, C. (2014) 'Data revolution: development's next frontier', World Politics Review 28 January.
Minkler, M. 'Enhancing data quality, relevance and use through community-based participatory research,' in Cytron, N, et al. (2014) What counts: harnessing data for America's future. With permission of the publishers, Federal Research Bank of San Francisco and the Urban Institute.
Miriri, D. and Blair, E. (2014) 'IMF says Kenya's new economic reforms are working. Reuters.
Natih, P. (2014) 'Multidimensional poverty in Indonesia: sparking agency based evaluation' (Master of Philosophy Thesis). Oxford University, Oxford.
Monga, V. (2014) 'The Big mystery: what's big data really worth? Wall Street Journal.
Mukherjee, S. (2013) 'Conceptualisation and classification of caste and tribe by the Census of India', Journal of the Anthropological Survey of India 62(2): 805-820.
Murray, C.J.L. et al. (2014) 'Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013', *The Lancet* 384(9947): 1005-70.
Naseem, S.M. (2012). A review of studies on poverty in Pakistan. Islamabad: Pakistan Institute of Development Economics (www.pide.org.pk/pdf/books/series-6.pdf).
NZDFF (New Zealand Data Futures Forum) (2014) Harnessing the economic and social power of data. NZDFF.
North, D., Gal, I. and Zewotir, T. (2014) 'Building capacity for developing statistical literacy in a developing country: lessons learned from an intervention', Statistics Education Research Journal 13(2): 15-27.
Olivia, S., Gibson, J., Brabyn, L. and Stichbury, G. (2014) 'Monitoring economic activity in Indonesia using night light detected from Space'. Paper Prepared for the 12th Indonesian Regional Science Association Conference, Makassar Indonesia 2-3 June 2014.
Open Society Foundations (2014) Ethnic origin and disability data collection in Europe: measuring inequality – combating discrimination. Brussels: OSF.
Over, M. (2014) Using 'value of information' concepts to prioritize the data revolution. Center for Global Development.
PARIS21 (2008-2015) 'Informing a data revolution'. Discussion Paper Series. PARIS21 (www.paris21.org/advocacy/informing-a-data-revolution).
PARIS21 (2015) 'A road map for a country-led data revolution'. PARIS21.
Patil, D. (2015) 'A memo to the American People from U.S. Chief Data Scientist Dr. DJ Patil', The White House Blog.
Press Information Bureau, Government of India (2013) 'The National Food Security Ordinance: Highlights', 5 July (http://pib.nic.in/newsite/erelease.aspx?relid=p\97050).
Puri, R. (2012) 'Reforming the public distribution system: lessons from Chhattisgarh', Economic and Political Weekly 47(5): 74-81.
PWC (2014) Deciding with data: how data-driven innovation is fuelling Australia's economic growth. Australia: PWC.
Ramadurai, C. (2014) 'Drones to aid conservation efforts in India', Green Futures Magazine.
Ranson, V. (2015) Fighting crime with better data. UK Government, Cabinet Office.
Rawlings, L. (2013) Who's afraid of administrative data? Why administrative data can be faster, cheaper and sometimes better. Washington DC: World Bank.

Recorder Report (2015) Criticality of credibility. Business Recorder: Pakistan (www.brecorder.com/editorial s/0:/1151827:criticality-of-credibility/); Abbasi, Z. and W. Lillah (2015). Inflation, poverty numbers: 'Erratic CPI basket may cast serious implications'. Business Recorder: Pakistan (www.brecorder.com/ business-a-economy/189/1150869/).
Reyes, C. (2014) 'Impact monitoring using CBMS, PPT' (www.3ieimpact.org/media/filer_ public/2014/09/18/30measuring_direct_and_indirect_ impacts_of_policy_and_program_interventions__reyes.pdf).
Robertson, A. (2015) Can online classrooms help the developing world catch up? The Verge.
Rodriguez-Takeuchi, L. (2014) The 'X-file' and the need for civil registration and vital statistics systems post-2015. London: Overseas Development Institute.
Rojas, G. (2015) 'Harnessing technology to streamline data collection' The DHS Program, USAID.
Ronsmans, C. and Graham, W. (2006) 'Maternal mortality: who, when, where and why', Lancet 368 (9542): 1189–1200.
Russell, N. (2014) 'UN awards big data project for confronting crazy weather in rice production' CIAT blog (www.ciatnews.cgiar.org/2014/09/02/un-awards-big-data-project-for-confronting-crazy-weather-in-rice-production).
Samman, E. and Roche, J. (2014) 'A data revolution to match the ambition of "leaving no one behind".' Post2015.org.
Samman, E. and Rodriguez-Takeuchi, L. (2013) Old age, disability and mental health: data issues for a post-2015 framework. London: Overseas Development Institute.
Sandefur, J. and Glassman, A. (2014) The political economy of bad data: evidence from African survey and administrative statistics. Washington DC: Center for Global Development.
Secretaría de Salud (2011) *Búsqueda Intencionada y Reclasificación de Muertes Maternas en México*. Mexico City: Secretaría de Salud.
Shaxson, L. (2005) 'Is your evidence robust enough? Questions for policy makers and practitioners', *Evidence and Policy* 1(1): 101-111.
Shaxson, L. (2014) Investing in evidence: lessons from the UK Department for Environment, Food and Rural Affairs. Knowledge Sector Initiative.
Snaing, Y. (2014) 'Burma's ethnic minorities decry census, jostle for advantage'. Yangon: The Irrawaddy.
SCUK (Statistics Commission UK) (2008) Report 38 Official Statistics: Value and Trust. London: UK Statistics Authority.
Statistics Netherlands (2012) 'The Dutch Virtual Census'. Video: http://youtu.be/SLpDkcyenf0.
Statistics Netherlands (2014) 'Statistics Netherlands: Dutch census saves time and money'. Press release, 19 November.
Stott, A. (2014) Open data for economic growth. Washington DC: World Bank.
Sullivan, K. (2015) How big data from space helps life on Earth. World Economic Forum.
Tode, F. (2014) 'UN discusses humanitarian uses of drones', Robohub.
UIS (UNESCO Institute for Statistics) and UNICEF (2015) Fixing the broken promise of Education for All: findings from the Global Initiative on Out-of-School Children. Montreal: UNESCO Institute for Statistics.
UN (1982) Model Life Tables for Developing Countries. New York: United Nations.
UN (1983) Manual X: Indirect Techniques for Demographic Estimation. New York: United Nations.
UNDESA (United Nations, Department of Economic and Social Affairs) Population Division (2013) World population prospects: the 2012 revision, key findings and advance tables. New York: UNDESA.
Understanding Children's Work Programme (2011) *Understanding Children's Work in Bangladesh*. Rome.
UNESCO (2010) Education for All monitoring report: reaching the marginalised. Oxford: UNESCO.
UNESCO (2014) 'Technical note prepared for the Education for All Global Monitoring Report: around 250 million children of primary school-age are not reaching a minimum learning standard'. New York: UNESCO.
UN Global Pulse (2014) Mining Indonesian Tweets to understand food price crises. Jakarta: UN Global Pulse.
UN Statistics Division (2013) 'Population and Housing Censuses' (http://unstats.un.org/unsd/demographic/ sources/census/alternativeCensusDesigns.htm).
Varian, H. (2014 revised) 'Big data: new tricks for econometrics', Journal of Economic Perspectives 28(2): 3-28.
Wanjiku, R. (2013) 'Google, Microsoft pilot TV white space projects in Kenya and South Africa'. Nairobi: PC World.
Watkins, K. and Alemayehu, W. (2012) Financing for a fairer, more prosperous Kenya: a review of the public spending challenges and options for selected arid and semi-arid counties. Brookings.
The White House (2014) The 90 Day Review for Big Data.
WHO, UNICEF, UNFPA, World Bank and United Nations Population Division (2014) Trends in maternal mortality: 1990 to 2013, estimates by WHO, UNICEF, UNFPA, The World Bank and the United Nations Population Division. Geneva: World Health Organization.
Wild, L. et al. (2015) CARE's experience with community score cards: what works and why? London: Overseas Development Institute.
Willetts, D. (2013) Eight great technologies. Policy Exchange.
Wilmoth, J. et al. (2012) 'A new method for deriving global estimates of maternal mortality', *Stat Politics Policy* 3(2): 2151-7509.
World Bank (2012) Information and communication for development 2012: maximising mobile. Washington DC: World Bank.
World Bank (2013) The World Bank policy on access to information. Washington DC: World Bank.
World Bank Group (2014) Voice Agency and Empowering women and girls for shared prosperity. Washington DC: World Bank Group

World Bank (2015a) A measured approach to ending poverty and boosting shared prosperity: concepts, data, and the twin goals. Washington DC: World Bank.
World Bank (2015b) *World Development Report 2015: Mind, society and behavior*. Washington DC: World Bank.
World Bank/WHO (2014) Global civil registration and vital statistics: scaling up investment plan 2015-2024. Washington DC: World Bank.
WEF (World Economic Forum) (2015) Data-driven development: pathways for progress. Geneva: World Economic Forum.
World Vision (2014) 'Connecting maternal, new born and child health with human resource in Uganda's health sector budget, Advocacy in Action. Nairobi: World Vision.
World Vision International (2014) More than numbers: why better data adds up to saving the lives of women and children. World Vision International.
Zucman, G. (2014) 'Taxing across borders: tracking personal wealth and corporate profits', *Journal of Economic Perspectives* 28(4): 121-148.

This is one of a series of Development Progress research reports.

Development Progress is a four-year research project which aims to better understand, measure and communicate progress in development. Building on an initial phase of research across 24 case studies, this second phase continues to examine progress across countries and within sectors, to provide evidence for what's worked and why over the past two decades.

This publication is based on research funded by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

ODI is the UK's leading independent think tank on international development and humanitarian issues. Further ODI materials are available at odi.org.uk

Development Progress

Overseas Development Institute
203 Blackfriars Road
London SE1 8NJ

Tel: +44 (0)20 7922 0300
Email: developmentprogress@odi.org.uk

facebook.com/developmentprogressproject
twitter.com/dev_progress

Designed by Soapbox, www.soapbox.co.uk

Figures by Steven Dickie

developmentprogress.org