

The Challenge Of No Data and Poor Data

Alonso Ortiz, Claire Melamed

Introduction

On September 2015 world leaders adopted the Agenda 2030, the most comprehensive and ambitious global framework for sustainable development. The agenda's most visible components are its 17 Sustainable Development Goals (SDGs) and its 169 associated targets. Attracting much less attention, the United Nations Statistical Commission agreed on March 2017 to a set of 230 indicators to monitor countries' SDG implementation over the next fifteen years. This is an equally ambitious (and complex) monitoring agenda that has shone a light on the fact that global, national, and sub-national data ecosystems are not equipped to the task.

There is a fundamental problem at the heart of sustainable development efforts - a problem of unreliable or non-existent data and the lack of skills and institutional frameworks to use it. This is a challenge that, superficially, appears to be technical (for example, the lack of sectoral data by gender, socio-economic status, and age), but is fundamentally one of political priorities and choices.

The data landscape has been transformed in the last decades, with the data deluge (Big data, Artificial Intelligence) and new groups of producers and consumers that have blurred roles and broadened definitions. These new changes, and the political focus they bring, are an opportunity to solve some old problems in the data ecosystem: new approaches and various encouraging examples are contributing to improve country data ecosystems, connect data users and producers, reimagine data to drive action, and to improve multi-stakeholder collaboration. Ultimately, enhancing and strengthening data ecosystems for sustainable development to respond to country priorities will have a direct impact on achieving the SDGs.

Article

On September 27, 2015, world leaders adopted Agenda 2030, a comprehensive and ambitious global framework for sustainable development. The agenda's most visible components are its 17 Sustainable Development Goals (SDGs) and its

169 associated targets. The 17 goals range from ending poverty in all its forms (Goal 1), to ensuring sustainable production and consumption patterns (Goal 12), and reducing inequalities, within and among countries.¹ While this exercise is a continuation of the Millennium Development Goals (MDGs), the current agenda added some new thinking to the process. For example, the MDGs narrowed the attention to developing countries and issues.²³ By contrast, Agenda 2030 and its SDGs are universal and require all stakeholders to contribute to its achievement.

The SDGs also added a new, needed emphasis on the use of data to achieve and monitor the goals. As many have argued, a serious challenge the MDGs faced was the lack of available, timely, and accurate data.⁴ The commitment to reduce absolute poverty was at the heart of the MDGs and, as such, “progress was measured by the proportion of the population escaping poverty ...”⁵ China and India were key to the significant progress of this piece of the agenda. Nonetheless, national aggregates “often masked considerable differences at the sub-national and individual level”, which means that there was enough data to estimate national averages, but not enough to count those left behind.⁶

While the MDGs era brought laudable achievements, such as a significant rise in the number of countries providing sufficient data to perform trend analysis for MDG health-related indicators (maternal health and malaria prevention), progress was uneven. In the case of birth and death registries (“information on births and deaths by age, sex and cause is the cornerstone of public health planning”⁷), there was almost no improvement in 15 years.⁸⁹¹⁰ “The SDG imperative to leave no one behind is a different dynamic and requires population data. It is not possible to know if people are missing if there is no data to show that they exist in the first place.”¹¹ These lessons did not go unlearned - and goal 17 in the SDGs includes a target (17.18) that aims to support developing countries’ capacity to increase significantly their production of “high-quality, timely, and reliable data.”¹²

For this reason, the United Nations Statistical Commission’s Interagency and Expert Group on SDG Indicators (IAEG-SDGs) developed a global indicator framework (currently composed of 232 indicators to be continuously reviewed) to monitor countries’ SDG implementation over the next fifteen years. On March 2017, the UN Statistical Commission agreed on a draft resolution on the framework and recommended its adoption to the UN Economic and Social Council (ECOSOC).¹³

UN Under-Secretary-General for the Department of Economic and Social Affairs, Mr. Wu Hongbo, rightly pointed out that global indicator framework is the not

the end of the story but rather the beginning of it.¹⁴ He acknowledges that “an unprecedented amount of data needs to be produced and analysed – and it is evident that this will pose a significant challenge for national statistical systems, in developing as well as developed countries.”¹⁵

Additionally, in order to achieve Agenda 2030’s commitment to leave no one behind, these massive quantities of data will need to be disaggregated by income, sex, age, race, ethnicity, migratory status, disability and geographic location or other characteristics. The SDG monitoring agenda exposes the reality that global, national, and sub-national data ecosystems are not equipped to the task.¹⁶

At a point in time where *artificial intelligence*, *machine learning*, *big data*, *data science*, and *data revolution*, amongst other terms, are being absorbed and adapted within the international development arena, it seems paradoxical that tracking 232 global indicators is a herculean task.

In fact, we have never had such a vast amount of data in our hands about every aspect of our lives. The phones in our pockets, the satellites that fly overhead, the cars we drive and the computers we use are all collecting data at a rate unimaginable to previous generations. The challenge is to turn this data into useful information. As Anne-Marie Slaughter put in a recent lecture:¹⁷

More is not better: that much data is useless, [we] might as well have none. [...] Looking for a needle in a haystack is now looking for a byte in a data archive.

New data sources can give us new insights, but only if accompanied with the right mix of tools and processes for filtering, managing, and analysing. What we have is an abundance of “noise”: meaningful data is scarce.

In reality, the deluge of new data does not eliminate the problem of poor data – it just puts it in a new light. As well as challenges in capacity and political support for traditional statistics, to fully take advantage of new opportunities and data sources, a new type of capacity is also needed – data science, programming, and visualisation. Moreover, a new set of skills – to link existing data sets to new data sets, to get the most value from both – lies at the heart of taking full advantage of the data revolution to achieve sustainable development.

This agenda was at the center of the first United Nations World Data Forum.¹⁸ In January 2017, 1,500 participants from over 100 countries headed down to Cape

Town, South Africa to engage “in almost 100 sessions, where innovative ideas were explored, pledges were made, and proposals for concrete data solutions were presented.”¹⁹

During the Forum, the United Nations Statistics Division unveiled the Cape Town Global Action Plan for Sustainable Development Data,²⁰ which proposes six strategic areas for improving capacity building to support countries in the implementation of the SDGs.²¹ It calls for investment in modernising national statistical systems, and for the involvement of non-governmental data producers and users to collaborate in the realisation of this agenda. The Action Plan confirms the official recognition of a new, more complex data ecosystem in which the roles and definitions of different actors are in flux, as well as the recognition that official and non-official sources of data need to work together towards improving data for sustainable development.

So, the question now is how to do that in a way that puts people, and their lives and aspirations, at the center of the story.

The Political Economy of Data

One clue as to how better share and use data comes from Kibera, Kenya. Kibera is a slum, the largest in Nairobi, Kenya (it houses 60 percent of the city’s population),²² and in the African continent. As Erica Hagen describes, in MIT’s *Innovations Journal*, “international audiences are led to believe that Kibera is a kind of hell on earth...” This is perhaps not entirely misleading, given that its population lives in extremely harsh conditions with serious sanitation, health, education, security, and labour deficiencies, with an added threat of complex land and security of tenure issues.

Most media coverage and development efforts do not speak to the residents of Kibera, or treats them as passive agents with regards to their home. “This leaves the population disempowered and unable to use information to solve problems that are endemic to Kibera, including poverty, poor health, and the lack of public facilities for water, electricity, and sewage. Thus they cannot meaningfully join the debate about how to improve Kibera.”

It was from this insight that the Map Kibera project was born. Erica Hagen and Mikel Maron envisioned it “to alter the existing power dynamic by shifting the information dynamic so that the residents could use localised information to influence policy and development.” Map Kibera is community information project that, initially, set

out to create an open digital map on the OpenStreetMap²³ platform to make the invisible visible, and for people in the community to describe what is important to them and how to improve it. This project has transitioned into a series of community projects, such as the Voice of Kibera news website,²⁴ that were “founded on the premise that the digital age makes it possible to bypass traditional gatekeepers to information and data—or to ignore them completely, thereby allowing citizens to create and use new information systems.” These efforts have paid off: mapping the more than 350 informal schools in the community helped uncover how the government’s promise of free education has not reached Kibera’s students. The Open Schools Kenya interactive database has been responsible for new public schools and for targeted support towards teacher training and other resources.²⁵

The Map Kibera project shows how better data, in the right hands, and used in the right way, can be transformational. Data can be used to describe and quantify people’s needs, and then be used by the government to help to meet those needs. In their raw form, data are numbers and binary code; when transformed, data are about people. Our use of statistics must be accompanied with empathy as much as rigour: data are about and for people.

A Complex Data Ecosystem

Bringing the world’s best data to bear on the world’s worst problems is literally impossible to do without finding new ways to bring together the diverse producers and users of data.

New collaborations are happening all over the world. For example Facebook, the Organisation for Economic Co-operation and Development (OECD), and the World Bank joined forces in February 2016 to deploy The Future of Business survey. This provides monthly information regarding business sentiment, drawn from the 60 million Small and Medium Enterprises (SMEs) across the world with an active Facebook business page. The survey currently operates in 33 countries and 140,000 Facebook Page owners have taken the survey.²⁶ These data are a gold mine for policymakers, researchers, and businesses looking to better understand business perceptions of current and future economic activity – critical, since 90 percent of businesses in all economies are SMEs and they are the major source of employment in many countries.²⁷ New data infrastructures – in this case the Facebook platform – can be leveraged to generate traditional types of data – in this case a survey – all through the power of collaboration.

Taking the example of Map Kibera or the Future of Business survey and turning them into established practices will not be easy. The challenges are technical – how to bridge the different methods, standards, and formats that define different data sets so they can be combined to create new insights. The challenges are political – how to make data that might come from many different providers, and that pose complex privacy and security issues, available for the common good without infringing on people’s rights. Finally, the challenges are financial – how to invest in the new systems and support the huge increase in human capacity that will be needed if governments, civil society, and the private sector are to truly take advantage of the opportunity that lies ahead.

However, the cost of not rising to this challenge is very real. Through the SDGs, the world has committed to enable 767 million people to escape extreme poverty,²⁸ and tackle humanity’s hardest challenge: preventing catastrophic climate change (Goal 13). To meet these challenges requires all our human ingenuity, and if we do so without using all the information that is now available, we are functioning with one hand tied behind our backs. The key to rising to this challenge is simple: collaboration, collaboration, collaboration.

We need collaborations that spur innovation by utilising existing sources of data, that leverage existing resources, that match data supply and demand, and set about smart funding. Such collaborations need to become the norm and not the exception. Most importantly, collaborations need to be built on the fact that data is about people and acknowledge that we know the least about the very people the SDGs claim to care the most: to leave no one behind we need to ensure that strong sub-national, national, and global data ecosystems are equipped to provide the right data, about the right people, at the right time.