Sabrina Juran, M.Sc, Technical Specialist, Data and Research, United Nations Population Fund (UNFPA) and Patrick Gerland, Ph.D., Chief, United Nations Department of Economic and Social Affairs (UNDESA), Population Division.

Mobilizing a Data Revolution for Sustainable Development

Access to and use of data forms the basis for effective decision-making. The international development community plays a pivotal role in the data revolution through support to countries in their efforts to create the evidence base for informed decision-making and policy formulation as well as to meet and monitor progress towards global development objectives.

In 2012, anticipating the conclusion of the Millennium Development Goals (MDGs) in 2015, the United Nations Secretary-General convened a High-Level Panel (HLP) of Eminent Persons on the Post-2015 Development Agenda. The report, "A New Global Partnership: Eradicate poverty and transform economies through sustainable development" called for a "data revolution for sustainable development" to improve the quality of statistics and information available to citizens" that can "actively take advantage of new technology, crowd sourcing, and improved connectivity to empower people with information on the progress towards the targets."[1]

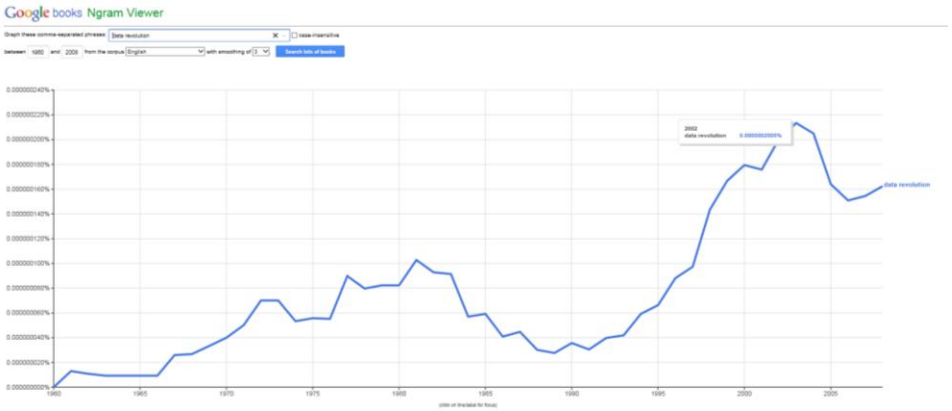
The report noted that the growth in the capabilities of information technology provides an opportunity to strengthen the use of evidence-based data and statistics for decision-making as well as for accountability purposes. It stressed the need for data to reach the most vulnerable populations. Better data are needed to support efforts to track progress on development initiatives, concluding. A "true data revolution would draw on existing and new sources of data

[1] United Nations. 2013: A New Global Partnership: Eradicate Poverty and Transform Economies through Sustainable Development. The Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda. http://www.un.org/sg/management/pdf/HLP_P2015_Report.pdf

to fully integrate statistics into decision making, promote open access to, and use of, data and ensure increased support for statistical systems."[2]

The idea of a Data Revolution is not new, and goes back at least since the invention of modern computers in the 1960s, but the idea has gained in popularity especially with the democratisation and expansion of the internet in the 1990s, mobile technology, remote sensing, Global Positioning System (GPS), geographic information system (GIS) and an increasing volume of digital information, especially sensor-based information becoming publicly available, as seen on Figure 1. The topic has even gained further in popularity in the last few years as seen on Figure 2 with a growing interest, new coverage and online searches about the concept.

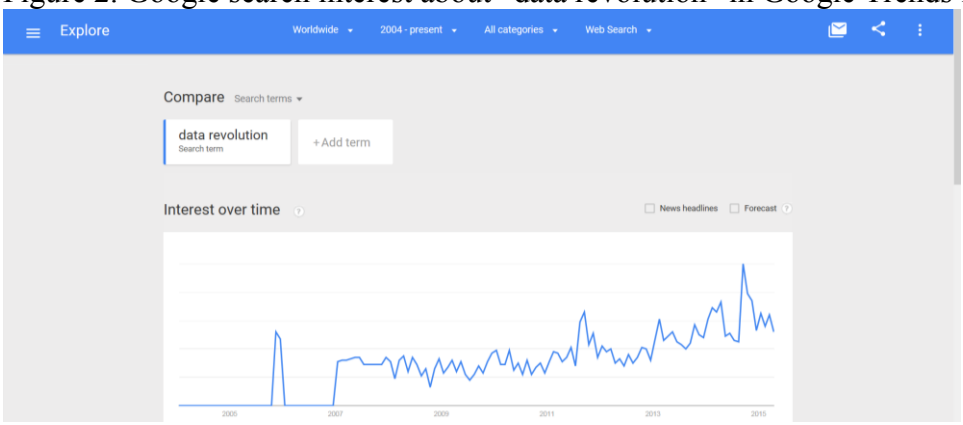
Figure 1. Frequencies for "data revolution" search terms in Google Books 1960-2008 Ngram Corpus



Source: Google books Ngram Viewer http://books.google.com/ngrams, English corpus, queried on 19 April 2015

[2] United Nations. 2013: A New Global Partnership: Eradicate Poverty and Transform Economies through Sustainable Development. The Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda. http://www.un.org/sg/management/pdf/HLP_P2015_Report.pdf

Figure 2. Google search interest about “data revolution” in Google Trends for 2004-2015



Source: Google Trends Viewer http://www.google.com/trends/, English corpus, queried on 19 April 2015

The 2013 HLP report set out the broad case for the data revolution. As such, the idea for a ‘data revolution for sustainable development’ was born. From the point of view of developing countries, the case for a revolution is very strong. Political attention generated by the launch of the post-2015 development agenda and the sustainable development goals (SDGs) present an opportunity to raise the profile of statistics and the systems that generate them. In the process leading up to the post-2015 development agenda, there has been ample opportunity to build on what has been achieved since the Millennium Development Goals were agreed upon. Lastly, by employing better data to strengthen and promote the voice of the poor, accountability can be increased and the effective use of resources ensured.

After the High-Level Panel highlighted the need for a data revolution as part of an accountability framework, in August 2014, the Secretary-General created an Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG) to shape an ambitious but realistic vision for a future development agenda beyond 2015 to succeed the Millennium Development Goals (MDGs) and to assess new opportunities linked to innovation, technical progress and the surge of new public and private data providers to support and complement conventional statistical systems and strengthen accountability at the global, regional and national levels.

The resulting report, "A World that Counts: Mobilising the Data Revolution for Sustainable Development,"[3] explored key opportunities and risks associated with a data revolution for sustainable development and presented recommendations for action, aimed to exploit these opportunities and mitigate the risks. The IEAG report defined the data revolution as "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 is 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"

Therefore, the data revolution entails many things including:

- Building on the wealth of official statistics that already exist in national statistical systems;
- Collecting more and better data on familiar as well as emerging aspects of well-being;
- Compiling data faster for real-time analysis and dissemination of data;
- Combining traditional data sources with new ones, including applications of "big data";
- Producing and using data in new ways to promote transparency and accountability;
- Enhancing the utilization of data, statistical literacy and improved quality of statistics to meet international standards;
- Using and integrating data in informed decision-making and policies that affect people's lives;
- Using data to monitor policy and programme implementation and assess results.[4]

The report of the Independent Expert Advisory Group on the Data Revolution underscored that the data revolution is not an end in itself, but a mechanism to improve the lives of people and to support equality and human rights and agreed that the data revolution is not about collecting or extracting more data, but about making an impact with existing and new data to improve social, economic, and environment conditions around the world.

[3] United Nations Data Revolution Group. 2014: Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development. http://www.undatarevolution.org/report/
[4] Sabrina Juran. 2014: Data for Sustainable Development. Available at: http://post2015.org/2014/01/30/data-revolution-for-sustainable-development/

The goals and targets of the Post-2015 Development Agenda must take into account current and future populations, changing age structures, mobility patterns, city growth and changes in the spatial distribution of people. Indicators, in turn, must be disaggregated by sex, age, social, economic, and ethnic characteristics. Good quality and timely data need to be analysed, disseminated and effectively used to ensure evidence-based decision-making. More data do not necessarily translate into better policies.

Improving the coverage, quality, timeliness, accessibility and completeness of national civil registration data is already acknowledged as an urgent priority across the developing world[5]. However, similar attention needs to be paid to improving the quality, coverage and local utilization of census, and survey data, but also service statistics from administrative systems, and the public and private sectors. National data archives and public use datasets needs to be made more widely and systematically available to make it easier for a wider range of data producers and users to preserve, to access and to use statistical data, both at the national[6] and international[7] levels. Support for analysis and the dissemination of data from the 2010 round of census and the upcoming 2020 round needs to be increased[8]. Not only will census data provide the baseline numbers for many indicators, including Sustainable Development Goal (SDG) indicators, but also the basis for drawing nationally representative sampling frames for surveys and other data collection activities, including big data. Surveys, such as the Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS), should include questions that are essential to understanding and measuring achievements for populations with unique needs, including young people and adolescents between the ages of 10 to 14 years[9]. Data gaps need to

[5] World Bank and WHO. 2014. Global Civil Registration and Vital Statistics. Scaling up Investment Plan 2015–2024. May 28, 2014. http://www.worldbank.org/en/topic/health/publication/global-civil-registration-vital-statistics-scaling-up-investment

[6] See the Accelerated Data Program (http://adp.ihsn.org/) by PARIS21 Secretariat at the Organisation for Economic Cooperation and Development (OECD) and the World Bank, and the International Household Survey Network (http://ihsn.org/) for a central inventory of censuses and surveys, guidelines, and open-source software for documentation, cataloguing, dissemination, and de-identification of microdata.

[7] See IPUMS-International for fully integrated and harmonized public-use census sample micro-datasets (https://international.ipums.org/international/), Terra Populus for census data integration with geospatial data and temporal comparability (http://www.terrapop.org/), and iDHS (https://www.idhsdata.org/) for the integration of a growing number of fully harmonized surveys to facilitate comparative research across time and countries.

[8] United Nations 2010 World Population and Housing Census Programme and revisions for the 2020 round of censuses: http://unstats.un.org/unsd/demographic/sources/census/

[9] See IHSN Question bank: http://www.ihsn.org/HOME/projects/survey-methods, IHSN Data Navigator http://datanavigator.ihsn.org/ and IPUMS-International (https://international.ipums.org/international-action/variables/group) to find topics and indicators collected in specific surveys or censuses

be closed to ensure that diverse groups of people and key issues become visible through disaggregation.[10]

The IEAG report identified two qualities that define the data revolution for sustainable development, namely the integration of new data with traditional data and an increase of the usefulness of data. Data integration, and especially geospatial and/or statistical linkage between census, survey data and other data sources (e.g., service statistics), is critical to use more effectively different types of information to produce small area estimations, and/or to estimate a richer set of indicators combining multiple sources of information, and to achieve greater levels of disaggregation than otherwise possible using independently each source of information. The discussions among members of the Independent Expert Advisory Group highlighted two big challenges for the current state of data; the challenge of invisibility, i.e. gaps in what we know from existing data and when we find out, i.e. timeliness or lag between initial data collection and public dissemination/use of information, and the challenge of inequality, i.e. gaps between those with and without information, and what they need to know to make their own decisions.

While the importance of monitoring progress toward achieving internationally agreed development goals has been recognized, available data are inadequate for this purpose, both for the MDGs and for the expanded set of SDGs that will define the post-2015 development agenda. Despite considerable progress in recent years, entire populations are not being counted and important aspects of people's lives and environmental conditions are still not measured at all or only too infrequently. This is because too often, existing data remain unused because they are released too late or not at all, not well documented and harmonized, or not available at the level of disaggregation needed for decision-making. As the international development community embarks on the Post-2015 Development Agenda with the Sustainable Development Goals (SDGs), tracking progress on new goals in new areas with even more indicators will increase the already high demands on National Statistical Offices to collect and analyse data.

Because of the importance of national statistical systems in collecting and compiling the data, the data revolution has to start there. There is no one size fits all data revolution.

[10] United Nations. 2015: Report of the Secretary-General. Monitoring of population programmes, focusing on integrating population issues into sustainable development, including in the post-2015 development agenda. E/CN.9/2015/4. http://daccess-ods.un.org/access.nsf/get?open&DS=E/CN.9/2015/4&Lang=E

Inequalities in availability of basic data about individuals, households and communities are the primary obstacle to reliable monitoring of development outcomes. More importantly, data must support national policy-formulation, monitor outcomes and help to empower poor people and communities at the local level. "Data are essential for good governance, decision making, policy formulation, strategic planning, resources allocation, etc. and demographic data lie at the heart of the monitoring system, as population numbers are the common denominator used in constructing all indicators of human development (income per capita, literacy rates, educational attainment, life expectancy, etc.). However, demographic data systems in many countries are inadequate in various respects."[11]

"People, and in particular the human condition, are at the core of the development agenda. To understand the situation today, we also need to know the past, and therefore we essentially depend on existing sources of information, and data already collected through traditional means. The Data Revolution offers great hopes and promises for faster, easier and better access to information, but ultimately we need a lot of information about people, and most of this information must be obtained with the informed consent of individuals and communities, and through their active collaboration in providing accurate information about their livelihood, living conditions, families, etc."[12]

The IEAG made specific recommendations on how to mobilise a data revolution for sustainable development and laid out four areas in which change is needed: principles and standards, capacity and resources, and partnerships and leadership as well as technology and innovation.

One proposal refers to the establishment of a process to create a "Global Consensus on Data" to adopt principles concerning legal, technical, privacy, geospatial and statistical standards which, among other things, will facilitate openness and information exchange and promote and protect human rights. If improvements in data are to lead to changes in peoples' lives, then data must be accessible and able to be used.

[11] United Nations Department of Economic and Social Affairs. Population Division. 2014. The Data Revolution for Development. What it means for the UN Population Division/DESA as users of nationally generated data. https://post2015.files.wordpress.com/2014/10/iaeg-dr_2014_unpd-final.pdf

[12] United Nations Department of Economic and Social Affairs. Population Division. 2014. The Data Revolution for Development. What it means for the UN Population Division/DESA as users of nationally generated data. https://post2015.files.wordpress.com/2014/10/iaeg-dr_2014_unpd-final.pdf

But, improving data is a development agenda in its own right. Existing gaps can only be overcome through the strengthening of capacities and new investments.

As the IEAG mentioned, 'strengthening national capacities will be the essential test of any data revolution'. Financial support could be used to support change at the national level, supporting dialogue between data providers and data users, enabling new and useful partnerships between public sector, private sector and civil society, investing in the technological infrastructure, and rewarding measurable improvements in the production and use of high quality data.

However, a new funding stream might be necessary to support the data revolution. Ensuring funding for investment in data will not be an easy task. Critically, the money must be spent in ways that enable and incentivise the changes that are needed to take advantage of the revolutionary possibilities in the data landscape.

The IEAG proposed a "Global Partnership for Sustainable Development Data" to mobilise and coordinate the actions and institutions required to make the data revolution serve sustainable development. Global level partnerships and leadership can help in at least four ways: (1) to consolidate and share emerging lessons and develop standards, (2) to broker necessary partnerships, (3) to develop regional and global technology infrastructure, and (4) to showcase best practice and encourage innovation. This too, needs resources, and political support to drive it. While the "World Forum on Sustainable Development Data" will bring together the whole data ecosystem to share ideas and experiences for data improvements, innovation, advocacy and technology transfer, the "Global Users Forum for Data for SDGs" will ensure feedback loops between data producers and users, help the international community to set priorities and assess results.

Most of the action, initiatives, and financing required to drive the data revolution will happen at the national and local levels. While official statistics will be at the core, ignoring the potential for innovation to solve problems, create new possibilities, and leapfrog over current technologies, will in the long run be a waste of resources. It is important that resources are available to ensure that there are incentives to innovate in the public interest as well as for the private sector. A starting point could to be to explore how new innovations could help to fill gaps

in data for the new Sustainable Development Goals, along the lines of the data labs proposed by the IEAG report.

A proposed global "Network of Data Innovation" could be created through which technology and innovation could be shared and used for the common good. Today, in the private sector, analysis of big data is commonplace – with consumer profiling, personalised services, and predictive analysis to optimise sales. Similar techniques could be adopted to gain real-time insights into people's wellbeing and to target aid interventions to vulnerable groups. Such innovations offer exciting new opportunities, but also throw up big challenges around privacy, public trust, and the potential abuses of data. Legal frameworks have not yet caught up with rapidly advancing technology.

In defining the Sustainable Development Goals to take us through to the year 2030, there is an opportunity to discover new ways of assessing wellbeing, measuring global development and making interventions in times of crisis. The United Nations Population Fund (UNFPA) and the United Nations Population Division believe that development takes place in locations, communities and countries where poor people live. To be effective the data revolution must not only deliver more data, but more importantly, these data must lead to more effective action on the ground. It thus means to get the right data to the right people at the right time and in the right format. The data revolution has to deliver better data and lead to better development outcomes in all countries. Development progress will depend on the reduction of inequalities both within and between countries. Therefore, effective monitoring of inequities and disparities for a wide range of demographic indicators and service statistics will require greater access to aggregate data at the country level as well as disaggregated data at the sub-national level. Because of the importance of national statistical systems in collecting and compiling these data, the data revolution needs to start there.

At the global level, quality and comparability of international estimates and post-censal projections of population, fertility, family planning, mortality, international migration and urbanization produced by the UN Population Division, and used by the whole international statistical system as the common denominator used in constructing all indicators of human development (income per capita, literacy rates, educational attainment, life expectancy, unmet

need in family planning, etc.), will benefit from increased coverage, quality, timeliness and open accessibility of national data.

National ownership of data is a prerequisite for development, and the data revolution provides a major new and critical opportunity to address statistical capacity-building, in partnership with the private sector and other non-State actors, for the generation and timely utilization of high-quality geospatial and population-related subnational, national, regional and global data in all countries.