Scientific Advisory Board
of the Secretary-General of the United Nations
hosted by the
United Nations Educational, Scientific and Cultural Organization

Comments by Members of the Scientific Advisory Board of the of the UN Secretary-General on the IEAG draft Report on Data Revolution

30 October 2014

The UN Secretary-General Ban Ki-moon has asked an Independent Expert Advisory Group (IEAG) to make concrete recommendations on bringing about a data revolution in sustainable development. The Group is part of the Secretary-General's efforts to prepare a Synthesis Report requested by UN Member States ahead of the intergovernmental negotiations leading to the adoption of the post-2015 agenda. The UN Secretary-General Ban Ki-moon has also asked that the members of his Scientific Advisory Board (SAB) engage in the elaboration of his Synthesis Report including its Data Revolution part.

After a quick examination– given the extreamly tight deadline - of the draft Report on Data Revolution released on 24 October 2014 by the IAEG, the SAB members would like to offer the following comments for further consideration in the elaboration of the final version of the Report:

1. The draft Report is comprehensive, forward looking, focused, cogent, reflecting the UN principles for sustainable development, and practical in its recommendations.

2. SAB Members highly appreciate that the notions of "equity in access and use of data" as well as "quality and integrity" have been given importance and are listed as key principles.

3. With regard to **equity in access and use of data**, it is noted that, although an important para on page 4 highlighhs that 'Above all, this should be a revolution for equity in access and use [of] data", the recommendations of the Report do not refer to this fundamental problem of equity. The SAB Members suggest therefore that specific recommendations be formulated in this respect and proposes that the following key global priorities be included:
   a. Avoiding a data divide between rich and poor countries;
   b. Harmonization and standardization of data platforms for increased accessibility and exchange;
   c. Capacity building nationally and regionally ;
   d. Identification of vastly experienced (e.g. India, Malaysia) but not necessarily "rich" countries to help less endowed countries;
   e. Establishment of training institutes, from technical training to graduate education at universities; the newly energized United Nations University system could perhaps play a leading role;
   f. Job creation as part of ramping up.

4. One general question to be addressed from the outset could be: what should be prioritized? What should a government, say in Sub-Saharan Africa, do first? Join international organizations that might be created or build local national capacity? Is there an opportunity to create jobs, especially for youth and young women? Such a prioritization could be pursued by the proposed "Global

Partnership for Sustainable Development Data" that would bring together all stakeholders. As one of the first steps, such a partnership could develop a strategy to expand the equal data accessibility.

5. Attention should also be provided to priorities for applications, perhaps foremost being the Civil Registration and Vital Statistics (CRVS) and collection of health related data in innovative ways e.g. the "million deaths" study. It might also be good to see examples of the "data revolution" in other areas (e.g. in the context of environmental conservation and biodiversity, such as the Global Biodiversity Information Facility) so that we can understand the complexity better, and test whether our assumptions are correct.

6. SAB Members would be happy to work with the IEAG to further identify such examples as well as gaps that need to be addressed.

7. With regard to ensuring, from the scientific point of view, the **quality and integrity of data** for political and economical decision-making and everyday use, great efforts should be made in developing standards and methods to analyse and evaluate the data quality before they are used. This is even more important for the information-poor and technically weak countries. To meet this challenge, ***international collaborations on data science and technoloty research are urgently required. This has not been clearly pointed out in the recommended actions in the part of technology, innovation and analysis (pp 20-21).***

8. Furthermore, to achieve high-quality decision-making, processes, products, and services are key to ensure success of data revolution for sustainable development. ***A key scientific component of data quality monitoring, which will allow the objectives to be accomplished, is an emphasis on multidisciplinary methods for data quality evaluation and improvement***. Total quality management tools, statistical quality control in particular, could be employed for auditing and improving data quality. Data mining algorithms, which were developed to explore interesting and important relationships among data, should be used for data quality assessment. Moreover software reliability, which is a major tool in data processing, could also be incorporated into data quality modeling. Related research plan should make extensive use of such multidisciplinary methods in order to develop a systematic data quality management methodology that is generic in the sense that it may be applied to a wide variety of business and manufacturing processes with minimal process-specific tailoring required by the user. Certain curricula could provide training on data quality modeling of information systems (from an information engineering perspective) and robust data profiling and optimization (from a management science perspective), as well as in the statistical fundamentals. ***Therefore, a well-established multidisplinary international research project should be initiated by and within the UN system***, which will help to significantly advance data management and decision-making methodology. Statistical quality control methods for detecting and correcting data errors when operative systems continuously generate real-time data traffics with uneven data feed speed and unexpected delays. Multidisciplinary research would provide opportunity for researchers from statistics, computer science, and industrial engineering to work together on an important area in information science and engineering. Thus, the success of such a project will contribute significantly to data-driven technologies for data intelligence.

9. There is a **need to ensure a much stronger symbiosis between the data collectors/providers and the data users**, both for decision-making and adaptive governance (a concept which should also be considered in the document), but also in the synthesis and analysis that can help decision-making. Although these aspects are addressed near the end of the Report through the idea of a global users forum, and in the final "quick wins" section with its analysis and "state of the world dashboard", the SAB felt that this was a pretty superficial treatment of a process (or many processes) that is at least as complex as the entire data collection and handling aspect itself and that therefore would deserve an equally thoughtful document.

10. Data and “big data” are indeed becoming important areas to focus on. But just as we have seen failures in “global health governance” with the Ebola epidemic, it can be expected that failures of “global data governance” will occur. It is therefore important to make provisions in the institutional arrangements, including at the UN level, to avoid such failures.

11. With respect to “big data”, what is really important and useful (as is made clear in the preamble) is the analytics and decision support deriving from it. Indeed the biggest challenges in not being drowned by a big data tsunami involve the ability to analyze data and to do so even when data are corrupted by noise, missing entries and in some cases tampering. Having such analytics is really critical to sustainable development. It is therefore proposed that Data Analytics be treated under a new item to be added to those on pages 18-20. Moreover, the term ‘data disaggregation’ (item 2) has a rather more narrow technical meaning when applied to sources of data such as energy consumption, water consumption, traffic information, etc. than the one used on page 18.

12. With regard to Data protection and privacy (item 7, page 19) the discussion is highy appreciated. Some additional points are suggested below:
    a. Audit mechanism: civil society institutions need to be brought in to provide oversight over the uses of big data and their usage (this provision is frequently referred to as “watching the data gatherers/analysts”).
    b. Cyber security attacks run the gamut from hacktivisits to anarchists and adequate resilience mechanisms need to be provided to deal with data integrity and “insider” attacks to enable reliable analytics for the data.
    c. Especially in the context of privacy of data gathering for sustainable development, a clear value proposition needs to be established about the utility of the data gathering for the parties whose data is being collected. Public good is certainly one use (line 652), but better yet are “private goods”. For instance data gathering about market prices of farm produce, or health data can be used to provide specific individual services for communities.

13. There is an error in the Box on pages 6 and 7: Figure 2: The bottom table, referring to the Cote d'Ivoire, is wrongly labeled "Population density per sq Km". Instead, it should be labeled "Prevalence of Malaria (%)". This is a nice figure, but its implications are not clear. It would be important to clarify how this data is helpful in fighting malaria. It does not seem to have any relationship to the interesting discussions about the malaria Mtrac in Uganda, or the epidemiological models developed at the University of Minnesota presented in the same box.

14. It is felt the "Data we want” section, especially on pp.14-15, is weak on its analysis of the lack of capacity in developing countries, although this is mentioned partially in the recommendations at least, which can be improved.

The above mentioned comments and remarks are supplemented by a series of proposed editorial changes that are submitted as ‘tracked changes’ in an Annex to this note.

ANNEX

A WORLD THAT COUNTS – MOBILISING THE DATA REVOLUTION FOR SUSTAINABLE DEVELOPMENT

1. A data revolution for sustainable development

1.1 What is the data revolution?

Data are the lifeblood of decision-making, and the raw material for accountability. Without data, we cannot know how many people are born and die; how many men and women still live in poverty; how many children need educating, and how many teachers to train or schools to build; the prevalence and incidence of diseases; if water is polluted or if the fish stocks in the ocean are dangerously low; how many adolescent girls are getting pregnant and what policies are effective in helping them; what companies are trading and whether demand for their product is expanding.

To know what we need to know involves a deliberate and systematic effort of finding out. It means seeking out high quality information that can be compared over time, between and within countries, and continuing to do so, year after year. It means careful planning, spending money on technical expertise, robust systems, and ever changing technologies. It means building public trust in the data, and expanding people's ability to use it, recombine it with other sources, share it to get different disciplines to formulate an interpretation of same technical facts .. .

Since 2000, the effort involved in monitoring the Millennium Development Goals (MDGs) has spurred increased investment in just these things, to improve data for monitoring and accountablity. As a result, we know much more now about the state of the world and, particularly, the poorest people in it. But despite this significant progress, huge data and knowledge gaps remain particularly in the LDC countries and/or in situations where data sharing for political reasons or otherwise are restrictive (eg water availability across political boundaries for same river-basin) about some of the biggest challenges we face, and these gaps limits governments' ability to act and to communicate honestly with the public. Months into the Ebola outbreak, for example, it was still hard to know how many people had died, or where. Using anonymised call data records from cellphone companies can be extremely useful in monitoring transport networks associated with Ebola and other contagious diseases or that spread thru transport driven vectors.

This in turn will require another significant increase in the information and most importantly knowledge to act that is available to governments, civil society, companies and international organisations to plan, monitor, preempt, manage and be held accountable for their actions.

Thanks to new technologies, the volume, level of detail, the diversity of data sources, the sourcing of garbage or exhaust data in comercial operations (eg. Microwave link data that can enable large scale rain measurement without weather stations) and speed of data available on societies, the economy and the environment is without precedent.

The data revolution is **a revolution of possibilities** – of new set of sensing technologies ~~technologies~~Technologies at very low cost (eg nanosat constellations with Earth Observation sensors that monitors the world's natural resources in near real time at a fraction of the cost of conventional technologies) , data production & integration, big data analysis that is machine-driven (machine learning and AI technologies) and Collaboration & dissemination systems and new resources opening up to produce more and better data, as well as expanding what can be done with data.

maximum effect. To fulfil this promise, it must be done in a way that adheres to the highest standards of honesty, respect of privacy, rigour and impartiality that have been developed over decades and centuries of academic research, statistical practice and political negotiation. It must also incorporate best understanding of both the public and private sectors to generate public goods in a way that a measurement is reproducible for enabling trust in the process.

It is also **a revolution of expectations** – of people demanding that these changes and innovations be used to enhance their control over their own lives and the decisions that affect them. Data is the bedrock of trust & accountability**.**

Without deliberate actions, the opportunities will be slower in coming and more unequally distributed when they arrive, and the risks will be greater.

~~106~~107 However, there is a case to be made that there are scenarios in the middle of the road where the value of the data and the information exceeds the risks justifying the controlled risks that should be taken. An example of this is the use of anonymised call data records to monitor location of affected citizens post a natural disaster as this technology can help focus post disaster efforts with a presence capability.