DRAFT

**CONTRIBUTING TO a toolkit for country data roadmaps**

In support of its objective to improve the quality, availability and accessibility of development data, the Global Partnership for Sustainable Development Data (GPSDD) has established a series of working groups. The Country Roadmaps Working Group specifically aims to assist countries to develop a comprehensive strategy for improving their data collection and statistical capacity. A key priority for the group is to develop a toolkit that enables countries to assess the quality of their current statistical instruments and data collection methods. This note summarizes work completed by the UN Sustainable Development Solutions Network to support the development of such a toolkit. Specifically, it proposes three tools: (1) an Essential Minimum Data Package that is needed to measure and implement the SDGs and (2) a Common Template to identify data gaps at the country level and (3) a series of questions that should help build fill this Common Template and build a country-level Roadmap.

The first tool is a matrix which identifies and tabulates internationally accepted data sources relevant for the broad range of development topics that are targeted by the SDGs, in addition to making note of any international standards and best practices associated with each statistical package. Several previous recommendations have informed this work (see Key Resources section), and this proposal has been complemented with more in-depth research into each data collection method.

The second tool is a Common Template that allows multi-sector stakeholders at the country level to identify the data gaps specific to each country.

The third tool is set of questions that should help countries’ national teams (statistical offices, working in partnership with the broader data community, including civil society and private companies and alongside the GPSDD), to assess their current data capacity, and determine areas where investments need to be made for a more efficient monitoring system, based on identified global standards.

**Essential Minimum Data Package (Annex 1)**

This exercise looks at seven key official statistical tools: **i) Census Data, ii) Household Surveys, iii) Agricultural Surveys, iv) Administrative Data, v) Economic and Fiscal Statistics, vi) Geospatial Data, and vii) Environmental Data**. These seven core areas were identified by reviewing 100 potential SDG indicators, by cataloging the type of data required to construct each indicator and the most frequently used methodology (more details on this can be found in the *Data for Development* report). In addition, this work identifies agencies with expertise on each methodology, and associated caveats with each data sources – like the desired level of disaggregation, suggested frequencies of collection, and conformity to international standards.

**Common Data Template (Annex 2)**

The Common Data Template is a tool to help the local country teams assess data gaps. It is built based on the Essential Minimum Data Package. The rows are exactly the same as the rows in the Essential Minimum Data Package, while the columns incorporate some of the key questions about each data source included in the questionnaire.

**Questionnaire (Annex 3)**

The questionnaire is a preliminary product, intended to help the national data teams (NSOs and a broad community of national data actors) to evaluate their preparedness for SDG monitoring, and to subsequently draft a data roadmap. It was created as a stepwise exercise that either countries themselves could conduct (self-evaluation) or consultants could use (external evaluation) to evaluate the current data availability for each data source in the Essential Package. Questions are deliberately simple. They guide the respondents through important criteria for robust datasets needed for monitoring the SDGs. Particular emphasis was paid to the frequency of data collection, alignment with international standard, sample representativeness, level of disaggregation, and geo-spatial attribution of data.

Once all the questions have been answered, countries should be able to identify gaps that need to be filled to conform to global best practices, thus enabling them to best determine where to invest to build capacity for better SDG monitoring. For example, if the question *“Did your country conduct a population and housing census in the past 20 years?”* receives an answer ‘YES’, but the following question *“how many censuses were conducted in the past 20 years?”* receives an answer ‘ONE’, then the countries are able to identify that more capacity needs to be built nationally to conduct population and housing census every 10 years (as per the guidelines in the data matrix). This questionnaire should be further refined in partnership with national statistical offices and other data stakeholders.

**Concluding Remarks and Recommendations**

Monitoring the SDG agenda would require substantive improvements in national statistical capacities. The necessity of stronger and more systematic collection of administrative data to improve government performance and encourage evidence-based decision making is unquestionable. Consistent and timely collection of quality data on the varied dimensions of sustainable development requires a thorough evaluation of the existing data systems in a country, and knowledge of globally accepted data collection best practices so that data gaps can be bridged through smart investments.

Although there are not many internationally accepted methodologies for Geospatial Data and Environmental Data, SDSN would like to take this opportunity to emphasize the importance of developing data systems and monitoring programs for these two, in light of the post-2015 development agenda. Additionally, SDSN would like to point out new innovations in data systems and data sources, that have emerged with advances in technology, such as satellite data, remote sensing, drones (UAVs), crowd-sourcing, social media data, mobile phone data, webdata and others, that offer opportunities to collect data in more efficient, more frequent and cost effective ways, to be used as compliments to official statistical data collection instruments, and as substitutes in extreme data constrained environments.

Lastly, further to what has been included in this toolkit, SDSN recommends, as a next step, including basic methodological guidelines in the toolkit. These would help evaluate qualitative weaknesses in existing data collection systems, which also need to be improved. For example, a census conducted by overburdened enumerators may not result in high quality data. Additionally, it is also important to have tools to help measure technical capacity to monitor the SDGs at the national level for countries who may not have (or have understaffed) statistical offices, GIS departments, or environmental offices. Having these would assist countries evaluate their current institutional setup and make improvements and additions.

**Key Resources**

Bubb, P.J., Butchart, S.H.M., Collen, B., Dublin, H., Kapos, V., Pollock, C., Stuart, S. N., Vié, J-C. (2009*). IUCN Red List Index - Guidance for National and Regional Use. Gland, Switzerland: IUCN.*

Data Revolution Group (2014), *A World That Counts: Mobilizing the Data Revolution for Sustainable Development.*

Food and Agriculture Organization of the United Nations (2015), *Guidelines on International Classifications for Agricultural Statistics.*

Food and Agriculture Organization of the United Nations (2015), *World Programme for the Census of Agriculture 2020.*

Global Strategy (2014), *Assessing Country Capacity to Produce Agricultural and Rural Statistics.*

International Conference on Agricultural Statistics VI (2013) – Proceedings, *Improving Statistics for Food Security, Sustainable Agriculture, and Rural Development. Linking Statistics with Decision Making.*

Paris21 (2013), *Towards a Post-2015 Framework that Counts: Developing National Statistical Capacity.*

Paris21 Task Force on Improved Statistical Support for Monitoring Development Goals, *Household Surveys and the Millennium Development Goals.*

Sustainable Development Solutions Network (2015), *Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development.*

Sustainable Development Solutions Network (2015), *Indicators and a Monitoring Framework for the Sustainable Development Goals: Launching a Data Revolution.*

The World Bank (2000), *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study.*

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2003), *Information Tools for the Preparation and Monitoring of Education Plans.*

United Nations Office for Disaster Risk Reduction (2013), *Global Assessment Report on Disaster Risk Reduction.*

United Nations Statistics Division – DESA (2005), *Household Sample Surveys in Developing and Transition Countries.*

United Nations Statistics Division – DESA (2008), *Principles and Recommendations for Population and Housing Censuses – Revision 2.*

United Nations Statistics Division – DESA (2010), *Manual on Statistics of International Trade in Services 2010.*

United Nations Statistics Division-DESA (2010), *Economic Census: Challenges and Good Practices.*

USAID (2006), *Guide to DHS Statistics: Demographic and Health Surveys Methodology.*

World Bank Group|World Health Organization (2014), *Global Civil Registration and Vital Statistics: Scaling up Investment Plan 2015-2024.*

World Health Organization (2004), *Developing Health Management Information Systems: A Practical Guide for Developing Countries*

**ANNEX I: ESSENTIAL MINIMUM DATA PACKAGE**

Please refer to the companion excel file, sheet 1.

**Additional Recommendations**

1. All statistical packages should have national coverage.
2. Modern technology (tablets or mobile phones) should be used for data collection.
3. All microdata should be made available in open formats after anonymization to protect personal privacy.
4. Survey and census metadata should conform to the DDI standard.
5. All data should be geo-located.

**ANNEX II: COMMON DATA TEMPLATE**

Please refer to the companion excel file, sheet 2.

**ANNEX III: QUESTIONNAIRE**

**Section 1: Census Data**

**Population and Housing Census (High Priority)**

* Did your country conduct a population and housing census in the past 20 years?
* How many censuses were conducted in the past 20 years?
* Were the Principles and Recommendations for Population and Housing Census, Rev. 2 followed for designing the questionnaire? (available @ http://unstats.un.org/unsd/demographic/sources/census/census3.htm)
* Were the census records geo-located? Did they include a postal code, administrative unit code, or other location identifier for each household surveyed?

**Economic Census (Medium Priority)**

* Did your country conduct an economic census in the past 10 years?
* Were the census records geo-located? Did they include a postal code, administrative unit code, or any other location identifier for each firm or enterprise surveyed?

**Agricultural and Livestock Census (Medium Priority)**

* Did your country conduct an agricultural and livestock census in the past 10 years?
* Were the FAO’s World Programme for the Census of Agriculture guidelines used for designing and conducting the census? (available @ <http://www.fao.org/economic/ess/ess-wca/en/> )
* Were the census records geo-located? Did they include a postal code, administrative unit code, or any other location identifier for each farm/agricultural enterprise surveyed?

**Section 2: Household Surveys**

**Demographic and Health Surveys (High Priority)**

* Does your country conduct sample surveys for demographic and health monitoring?
* How many have been done over the past 10 years? What was average time lag between two consecutive surveys?
* Was the DHS Program survey employed or was it your own design? (available @ <http://www.dhsprogram.com/publications/publication-dhsq7-dhs-questionnaires-and-manuals.cfm> )
* Was the survey gender disaggregated?
* Were the survey records geo-located? Did they include a postal code, administrative unit code, or any other location identifier for each individual/household surveyed?
* What is the smallest subnational unit for which reliable results were obtained?

**Health and Education Surveys (High Priority)**

* Does your country conduct surveys on education and health?
* How many have been done over the past 10 years? What was the average time lag between two consecutive surveys?
* Did you conduct UNICEF’s MICS survey or was it your own design? (available @ <http://www.ceecis.org/mics/printed_material/User_Guide_to_MICS_eng.pdf> )
* Was the survey gender disaggregated?
* Were the survey records geo-located? Did they include a postal code, administrative unit code, or other location identifier for each individual/household surveyed?
* What is the smallest subnational unit for which reliable results were obtained?

**Income and Expenditure Surveys (High Priority)**

* Does your country conduct surveys on income and expenditures or household consumption, employment, wages, community facilities and access, for measuring living standards?
* Did you conduct The World Bank’s LSMS survey or was it your own design? (available @ <http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,contentMDK:21555942~pagePK:64168445~piPK:64168309~theSitePK:3358997,00.html> )
* How many have been done over the past 10 years? What was the average time lag between two consecutive surveys?
* Was the survey gender disaggregated?
* Were the survey records geo-located? Did they have a postal code, administrative unit code or any other location identifier recorded for each individual/household surveyed?
* What is the smallest subnational unit for which reliable results were obtained?

**Other Survey Programs (Medium Priority)**

* Does your country do any other household surveys? What were the purpose of these other surveys?
* How frequently have they been conducted? Is the frequency of data collection greater than 5 years?
* Are the surveys gender disaggregated?
* Are the survey records geo-located? Did they have a postal code, administrative unit code or any other location identifier recorded for each individual/household surveyed?
* What is the smallest subnational unit for which reliable results were obtained?

**Section 3: Agricultural Surveys (Medium Priority)**

* Has your country conducted surveys to monitor agricultural land holding, productivity/yields, cost of cultivation, cost of production, input use?
* Have you followed the FAO guidelines to design your agricultural data collection programs? (available @ <http://www.fao.org/economic/the-statistics-division-ess/world-census-of-agriculture/conducting-of-agricultural-censuses-and-surveys/chapter-6-the-survey-design/en/>)
* Does your country maintain a comprehensive agricultural land records system?

**Section 4: Administrative Data (High Priority)**

**Education Management Information Systems (EMIS)**

* Does your country have an education management information system to maintain disaggregated records of education related data at the national level?
* If yes, have you evaluated the national EMIS to check compatibility with the UNESCO framework for EMIS? (available @ <http://unesdoc.unesco.org/images/0015/001568/156818eo.pdf> )
* Can the collected information be disaggregated by gender and at the lowest level of administration in your country?
* Is the data available in an open format and, if so, where?

**Civil Registration and Vital Statistics (CRVS)**

* Does your country have a Civil Registration and Vital Statistics Management system?
* Have you compared the system to the WHO framework for CRVS? (resources available @ <http://unstats.un.org/unsd/Demographic/crvs/globalcrvs.html> )
* Can the collected information be disaggregated by gender and at the lowest level of administration in your country?
* Is the data available in an open format and, if so, where?

**Management Information Systems for Health (MIS)**

* Does your country have a Health Management Information system?
* Have you compared the system to the WHO framework for Health MIS? (available @ <http://www.who.int/healthmetrics/documents/Components_of_a_strong_HIS.pdf> )
* Can the collected information be disaggregated by gender and at the lowest level of administration in your country?
* Is the data available in an open format and, if so, where?

**Section 5: Economic and Fiscal Statistics (High Priority)**

* Does your country collect labor force statistics regularly? How many labor force surveys were done in the past year? Do the surveys collect data that is classified according to the International Standard Classification of Occupations (ISCO 2008)? If no, is there a concordance table that is generated by the national statistical office?
* Does your country conduct establishment surveys? How many were done in the past 5 years? Does the survey include a labor demand module? Does the collected data conform to the New Industrial and Business Statistics Program (NIBSP)? If not, is there a concordance between the local classification and international standard industrial classification (ISIC) issued?
* Does your country regularly collect and report price statistics, national accounts, macro-economic data, central bank and depository corporation data, interest rates to the IMFs e-GGDS? Do these data adhere to the IMF’s Special Data Dissemination Standards (SDDS), and the System of National Accounts standards (SNA 2008)?
* Does your country keep track of Trade Statistics, and Service and Tourism Trade Statistics? Are these databases conforming to SNA 2008 and Balance of Payments and International Investment Position Manual (BPM 6)?
* Are these data available in an open format and, if so, where?

**Section 6: Geospatial Data (Medium Priority)**

* Does your country maintain a National Infrastructure and Facilities Inventory (NIFI)?
* Does your country monitor land cover and land use periodically for changes, and calculation of other indicators?
* Does the national statistical office have adequate workforce to support geospatial analyses (GIS analysts etc.)?
* Does your country utilize satellite imagery and remote sensing analysis techniques as data sources?
* Does your country maintain a geospatial data portal which offers administrative and enumeration boundaries, physical features data, and other auxiliary geospatial information?
* Is the data available in an open format and, if so, where?

**Section 7: Environmental Data (High Priority)**

* Does your country monitor greenhouse gas emissions in accordance with UNFCCC guidelines for all land uses? Is this data disaggregated by industry?
* Does your country monitor freshwater access and water quality? Are DHS and MICS used to collect this data? What is the level of disaggregation at which this data is available? Is this disaggregated by gender? Are you aware of WHO and UNICEF’s Joint Monitoring Program (JMP)? What is the frequency of collection of this data?
* Does your country monitor urban and rural waste generation?
* Does your country collect data on cooking fuel usage in households? Are MICS and DHS used to collect this data? Is this data representative at the National, Rural/Urban levels? What is the frequency of collection of this data?
* Are other environmental factors like climate, air quality and biodiversity being monitored by your country? Have you utilized international resources like IUCN for biodiversity, CIESIN and NOAA for climate monitoring?
* Are you making efforts to monitor air quality at the local level on a daily basis by installing cheap equipment that is recommended by WHO?

**Data Innovations**

* Does your country have an open data portal?
* If so, which of the data mentioned above can be found there?
* Is your country ranked in either the Open Data Index or the Open Data Barometer (or both) and, if so, what is your country’s ranking?
* Is your country using any Big Data sources (satellite imagery, unmanned aerial vehicles, smart-meters, mobile phones, social media data, web data and similar) to lower costs of data production or as compliments or alternatives to traditional official statistical data?
* Is your country using tablets or smart phones or similar devices to collect data?