Briefing Notes on key terms for the National Data Roadmap Forum

Strengthening Censuses and Surveys

Utilising Administrative Data

Data Disaggregation

Exploring new types of data

Geospatial Data

Open Data

Citizen Generated Data

Big Data

Data Communication

Managing Data Ecosystems

Evidence-based Decision-making

Funding and Capacity Building

Introduction

The following briefing notes were compiled to ensure that all participants in the upcoming National Data Roadmap Forum are familiar with new key terms that will be explored during the 2-day event. Please see the references section for further information on each of the terms.

Below are short explanatory notes for the following terms:

- Open Data
- Big Data
- Geospatial Data
- Citizen Generated Data
- Data Ecosystem

Open data

What is Open Data?

Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike.

To give further details, open data refers to:

- **Availability and Access:** the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- **Re-use and Redistribution:** the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.
- **Universal Participation:** everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

There are 6 principles of open data

1. Open by default
2. Timely and comprehensive
3. Accessible and usable
4. Comparable and interoperable
5. For improved governance and citizen engagement
6. For inclusive development and innovation

This means that data must be freely available to users to access in a way that they can easily reuse it. For example, providing data in PDF format or in hard copy makes it difficult for analysts to reuse this data without first doing a considerable amount of work.

Open Data is still subject to all of the anonymization and data protection stipulations of official statistics. Thus not all data is eligible to become open data.

How does Open Data relate to SDG indicator production?

Opening Data is key to ensuring transparency in the effort to both achieve the SDGs and the methods used to produce data for SDG indicators. Open data allows for stakeholders to hold decision-makers to account with regard to progress against SDG targets and goals.

Opening official data means that it will be available in an easy to use format for any user to download and use for their purposes. This also has additional benefits in allowing a producer-user dialogue and feedback loops to improve the quality of data and can inform discussions about data harmonization to guard against conflicting estimates.

Ultimately the Open Data movement views data as a public good which citizens pay for through taxation and therefore should be available to all.

What activities relating to Open Data have been undertaken in Ghana to date?

In 2012 Ghana launched the Ghana Open Data Initiative including an online platform to act as a repository of micro datasets. This can be accessed here: http://data.gov.gh/. Open Data Watch is an organisation which conducts an inventory of openness with regard to country data. In 2016, Ghana ranked 137th in the World (http://odin.opendatawatch.com/Report/countryProfile/GHA?appConfigId=3) on the Open Data Inventory with particular issues in the fields of environmental and economic statistics. It also noted that coverage with regard to regional and district level data was of particular concern as were the download options offered, availability of machine readable formats and ability of users to access data for unrestricted use.

Plans are now underway at GSS to build an online platform focused on data for SDG indicators, as well as national and regional indicators that would allow users to interact with the data to track progress over time.

Big Data

What is Big Data?

Big Data is an umbrella term referring to the large amounts of digital data continually generated by the global population. The speed and frequency by which data is produced and collected—by an increasing number of sources—is responsible for today's data deluge. A large share of this output is "data exhaust," or records generated as a by-product of everyday interactions with digital products or services.

The private sector—including mobile phone carriers, credit card companies and social media networking sites—manages enormous data sets that hold rich insights. Companies analyze this data to support decision-making or provide market intelligence. More recently, public sector institutions have begun leveraging similar techniques to generate actionable insights for policymakers.

Big Data is characterized by the "3 Vs:" greater volume, more variety, and a higher rate of velocity. A fourth V, for value, can account for the potential of Big Data to be utilized for development.

Big Data is different from Open Data. Open Data refers to data that is free from copyright and can be shared in the public domain, amongst other characteristics (see above). That is not a defining characteristic of Big Data, which can be privately owned or have varying levels of access control.

How does Big Data relate to SDG indicator production?

Big Data is a relatively new source of data and pilot projects are ongoing regarding the use of Big Data for SDGs indicator production. What is clear is that Big Data is usually best used as a complementary source in combination with traditional sources of data such as censuses and surveys and administrative data sources which can be used to validate Big Data sources.

The benefits of Big Data for SDG indicator production are as follows:

- Big Data is continuously produced and if real time analysis is adopted can offer more timely estimates that censuses/ surveys
- Big Data is passively produced which means that there is no direct cost associated with the production of the data, though there may be cost associated with making data available in a particular format

To see a list of Big Data projects conducted by the UN's Global Pulse, click here http://www.unglobalpulse.org/projects.

What activities relating to Big Data have been undertaken in Ghana to date?

A project undertaken by UN Global Pulse in Ghana in relation to governance and peace could contribute to monitoring of goal 16 of the SDGs. For more information click here http://gappoll.unglobalpulse.net/ghana/.

Geospatial Data

What is Geospatial Data?

Geospatial data, GIS data or geodata has explicit geographic positioning information included within it, such as a road network from a GIS, or a geo-referenced satellite image. Geospatial data may include attribute data that describes the features found in the dataset.

Thus, geospatial data is that which has some sort of a location marker in it. When combined with demographic or socio-economic data, this can help to further disaggregate the information available.

How does Geospatial Data relate to SDG indicator production?

Geospatial data can significantly reduce the costs of monitoring the achievement of the SDGs and the associated targets and indicators. In particular, it can contribute significantly to the production of environmental statistics.

One of the key areas of geospatial data is Earth Observation (EO) data. EO data and information, which include satellite, airborne, land and marine-based data, play an essential role in underpinning the environmental dimension of the SDGs. Combined with demographic, statistical and other data, EO can be used for action and data-driven decision-making across government. EO can track global change in high resolution and in real time and has the potential to provide more dynamic disaggregated data to achieve the SDGs.

Geospatial data can also be used in the production of other types of indicators, for example in monitoring crop conditions in countries at risk of food insecurity, to measure air quality in cities, tracking forest loss etc. In short, it back be used not only to track progress across geographic locations but also across socio-economic groups and other levels of disaggregation.

What activities relating to Geospatial Data have been undertaken in Ghana to date?

There have been a number of activities around geospatial data involving government, academia, civil society, and the private sector in recent years. Ghana Statistical Service also has a Geographic Information Systems (GIS) section which is available to stakeholders to produce maps down to enumeration area level. Moreover, Ghana has published a national and sub-national spatial development plans.

However, there is much more potential for geospatial data to be explored for SDG indicator production.

Citizen Generated Data

What is Citizen Generated Data?

Citizen-generated data (CGD) is data that people or their organisations produce to directly monitor, demand or drive change on issues that affect them. This can be produced through crowdsourcing mechanisms or citizen reporting initiatives, often organized and managed by civil society groups. This is distinct from "big data" or social media data, which is indirectly created by citizens through interaction with media platforms.

There is much enthusiasm about the potential of CGD to raise citizens' voices and to contribute to the "data revolution", but can also be criticized for its lack of representivity or statistical rigor.

How does Citizen Generated Data relate to SDG indicator production?

Citizen generated data can contribute additional information to that produced by traditional data sources by elevating the voices of marginalized communities. This data can be produced about a particular topic in a short space of time, or can be generated by the public over a long period. Integrating citizen generated data can also increase accountability and transparency of efforts to achieve the SDGs.

Thus, citizen-generated data can:

- Give greater coverage
- Enhance accessibility
- Improve data quality
- Contribute qualitative data for indicator production

However, CGD sets are often vulnerable to critiques regarding the rigor of their data collection methodology, especially sampling methods. These concerns are at times valid, and can reflect a general lack of statistical expertise among civil society organizations. If Ghana were to prioritise citizen-generated data it would need to work closely with CSOs to mitigate this risk.

What activities relating to Citizen Generated Data have been undertaken in Ghana to date?

To date there have been few concerted efforts to integrate CGD with official statistics that GSS is aware of.
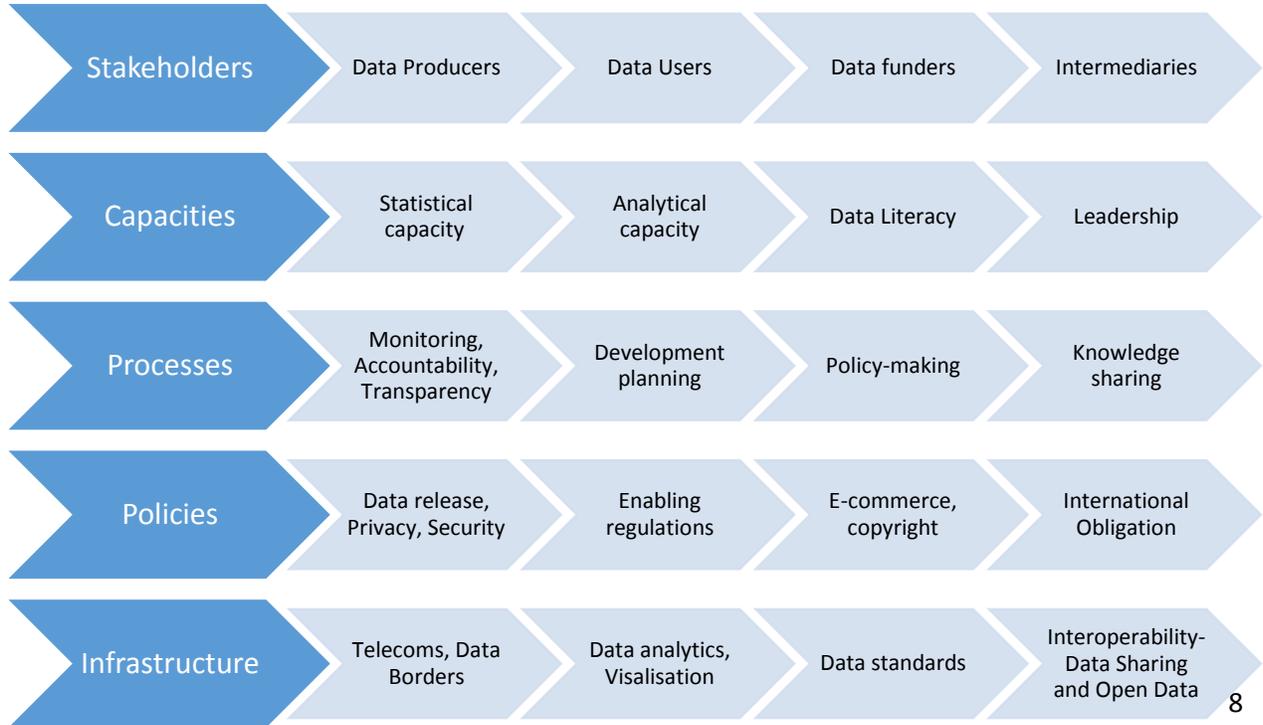
Data Ecosystem

What is a Data Ecosystem?

The term "Data Ecosystem" refers to the expansion in the actors and stakeholders interested in data production and use beyond the traditional data system. Previously, with regard to official statistics, we thought of the National Statistical Office (NSO) and the key Ministries, Departments and Agencies that produced data as comprising the National Statistical System (NSS).

In recent years however, the demand for data has increased exponentially. So too has the amount of data available and the sources of data, thanks to new technology. Today, it would not be possible for the traditional official statistics NSS to produce all of the data a country requires alone. It has become necessary to think about the wider system of actors who must be involved in data production, for example private sector organisations now have data which can be useful for official statistics such as mobile phone data.

This increase in data production is part of the data revolution that is currently going on across the world. The data revolution also refers to the increase in the number of users and uses of data. We must now think beyond traditional users of data, such as policy-makers and academia, to include civil society actors, private sector actors, the media etc.

It is the combination of these traditional and more recent actors, institutions and stakeholders in the production and use of data that is termed the "Data Ecosystem".

The elements of a data ecosystem are depicted below. In addition to an expansion of the stakeholders involved, there are also new capacities, processes, policies and infrastructure required to support the development of a thriving data ecosystem.

Stakeholders	Data Producers	Data Users	Data funders	Intermediaries
Capacities	Statistical capacity	Analytical capacity	Data Literacy	Leadership
Processes	Monitoring, Accountability, Transparency	Development planning	Policy-making	Knowledge sharing
Policies	Data release, Privacy, Security	Enabling regulations	E-commerce, copyright	International Obligation
Infrastructure	Telecoms, Data Borders	Data analytics, Visalisation	Data standards	Interoperability- Data Sharing and Open Data

References

Open Data

Open Data Charter, http://opendatacharter.net/principles/.

Open Knowledge International, *Open Data Handbook,* available at: http://opendatahandbook.org/guide/en/what-is-open-data/.

The Open Data Institute, *Guides to Open Data,* available at: http://theodi.org/guides.

Open Data Watch, *Open Data Inventory, Ghana,* available at: http://odin.opendatawatch.com/Report/countryProfile/GHA?appConfigId=3.

Big Data

UN Global Pulse, *Big Data for Development*, available at: http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf.

Big Data UN Global Working Group, available at: https://unstats.un.org/bigdata/taskteams/sdgs/.

Geospatial Data

Global Partnership for Sustainable Development Data, *Earth Observation Data to Support the SDGs,* available at: http://www.data4sdgs.org/earth-observation-data-to-support-the-sdgs/.

Group on Earth Observations, *Earth Observations and Geospatial Information: Supporting Official Statistics in Monitoring the SDGs,* available at: https://www.earthobservations.org/documents/meetings/201603_eo_sdgs_ny/2016_geo_un_flyer.pdf.

Citizen Generated Data

Civicus, *Citizen generated data and governments,* available at: http://civicus.org/images/citizen-generated%20data%20and%20governments.pdf.

United Nations, *Everyone Counts: Using citizen generated data to monitor progress against the SDGs,* available at: https://sustainabledevelopment.un.org/partnership/?p=11910.

Global Partnership for Sustainable Development Data, *Making Use of Citizen-Generated Data,* avalabel at: http://www.data4sdgs.org/guide-making-use-of-citizen-generated-data/.