COUNTING ON THE WORLD

Building Modern Data Systems for Sustainable Development

SUSTAINABLE DEVELOPMENT SOLUTIONS NETWORK
A GLOBAL INITIATIVE FOR THE UNITED NATIONS

TRENDS
Thematic Research Network on Data and Statistics

Members of the Sustainable Development Solutions Network **Thematic Research Network on Data and Statistics (SDSN TReNDS)**, include:

CO-CHAIRS

Enrico Giovannini
University of Rome Tor Vergata, Rome, Italy

Robert S. Chen
Center for International Earth Science Information Network (CIESIN), New York, U.S.A.

MEMBERS

Barbara Ryan
Group on Earth Observations (GEO), Geneva, Switzerland

Bram Govaerts
International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

Calogero Carletto
The World Bank, Washington, D.C., U.S.A.

Chukwudozie Ezigbalike
United Nations Economic Commission for Africa, Addis Ababa, Ethiopia

Eduardo Sojo Garza Aldape
National Laboratory on Public Policy – CIDE – Mexico City, Mexico

Emmanuel Letouzé
DataPop Alliance, New York, U.S.A.

Eric Swanson
Open Data Watch, Washington, D.C., U.S.A.

Gilberto Câmara
National Institute for Space Research (INPE), São José dos Campos, Brazil

Guo Huadong
Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China

Kate Crawford
Microsoft Research & MIT Center for Civic Media, New York, U.S.A.

Keith Shepherd
World Agroforestry Center (ICRAF), Nairobi, Kenya

Laveesh Bhandari
Indicus Foundation, New Delhi, India

Shaida Badiee
Open Data Watch, Washington, D.C., U.S.A.

Molly Jahn
The University of Wisconsin – Madison, Wisconsin, U.S.A.

Nicolas de Cordes
Orange, Paris, France

Philipp Schönrock
Centro de Pensamiento Estratégico Internacional (CEPEI), Bogotá, Colombia

Sabina Alkire
Oxford Poverty and Human Development Initiative (OPHI), Oxford, U.K.

Tom A. Moultrie
Centre for Actuarial Research, University of Cape Town, Cape Town, South Africa & International Union for Scientific Study of Population

Virginia Murray
Public Health England, London, U.K. and United Nations Office for Disaster Risk Reduction (UNISDR) Scientific and Technical Advisory Group member and vice-chair (2008-2017)

William Hoffman
World Economic Forum, New York, U.S.A,

Alex Fischer (ex officio member)
Smith School of Enterprise and the Environment, University of Oxford, Oxford, U.K.

Marc A. Levy (ex officio member)
Center for International Earth Science Information Network (CIESIN), New York, U.S.A.

THE NETWORK IS SUPPORTED BY:

Jessica Espey
Senior Adviser, SDSN

Melika Edquist
Manager, SDSN

ACKNOWLEDGEMENTS

Jessica Espey wrote this report with input from members of SDSN TReNDS including Barbara Ryan, Bram Govaerts, Calogero Carletto, Chukwudozie Ezigbalike, Eduardo Sojo Garza Aldape, Emmanuel Letouzé, Enrico Giovannini, Eric Swanson, Laveesh Bhandari, Molly Jahn, Nicolas de Cordes, Philipp Schönrock, Robert S. Chen, Sabina Alkire, Shaida Badiee, Tom A. Moultrie, Virginia Murray, Alex Fischer and Marc A. Levy. Specific text boxes were provided by Alex Fischer, Tom Moultrie, Virginia Murray, Marc Levy, Eric Swanson and Shaida Badiee, as well as Fredy Rodriguez (CEPEI) and Dierdre Appel (Open Data Watch).

The group would like to acknowledge and thank the following people for their thoughtful inputs: Melika Edquist and Guido Schmidt-Traub (SDSN), Alison Holder (Equal Measures), William Sonntag and Steven Ramage (Global Earth Observation Secretariat), and Sarah Lucas (Hewlett Foundation).

We are particularly grateful to the Hewlett Foundation for their generous support for the work of SDSN TReNDS, and to the Federal Government of Germany for their support for select, data-focused Solutions Initiatives (SDSN projects which aim to promote new technologies, business models, institutional mechanisms, policies and combinations thereof that can dramatically accelerate progress towards sustainable development).

The title of this report was inspired by the UN Secretary General's Independent Expert Advisory Group on the Data Revolution. Their report, *A World That Counts*, was published in 2015. It provided a comprehensive assessment of the state of current data and information systems and the potential offered by the 'data revolution' for the monitoring and achievement of sustainable development. Five of the original independent expert group are members of SDSN TReNDS and it is our intention to build upon that seminal work, providing a more up to date, practical pathway to achieve modern data systems that integrate the most promising aspects of the data revolution for sustainable development.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

"...Near enough to every one of us is in the middle of this data thing. Today, just using a supermarket or getting a flight is enough, even if you don't use the internet and still have a pedal wireless at home. It's a big shift that's going on, possibly the biggest structural shift in the economy in a generation. And it has a long way to run yet, by all the evidence."

Peter Harris AO, Chairman of the Productivity Commission of the Australian Government[1]

The term 'data revolution' is ubiquitous. Over the last five years, we have heard time and time again about the potential of new technologies and big data to transform the way we do business within both the private and public sectors. But the need for new approaches was brought into sharp relief during the negotiations on the new global sustainable development agenda. In 2015, 193 heads of state and government agreed to set the world on a path to a more sustainable future, through the pursuit of 17 Sustainable Development Goals (SDGs) of the 2030 Agenda for Sustainable Development ("2030 Agenda")[2]. These goals are predominantly quantitative: aiming to eradicate poverty and forms of deprivation, prevent environmental degradation and level inequalities by 2030. They are backed up by a comprehensive set of targets and indicators that rely upon access to high-quality, granular data.

Unfortunately, much of the data required to monitor the SDGs is unavailable. Issues relating to quality, timeliness, human and financial capacity, and lack of standardized methodologies all hamper our ability to comprehensively track this important agenda. Financial investment in statistical systems is urgently required to help rectify these problems, but we also need to harness the so-called 'data revolution,' bringing in private companies and other data innovators from academia, civil society and multilateral institutions to develop new technologies and approaches to monitoring sustainable development.

But data is not just required for monitoring. Achieving the ambitious SDGs also requires an evidence-based approach to governance. National and local leaders need to utilize data to help inform their planning, decision-making, and program design. They need data on the here and now, on the quality of services, on economic opportunities, and on the wellbeing of their population. But they also need historical data and forward-looking, modeled data to understand where we are coming from, understand trends over time, and help us prepare for the future and the seismic sustainable development challenges it brings with it.

This report lays out the necessary functions of modern statistical systems equipped to support sustainable development. Given the scale of the challenge, it proposes a multi-stakeholder approach in which private companies, academia, multilateral institutions and civil society support governments with the production, cleaning, compilation, dissemination and analysis of data. It identifies the range of actors that should be included and their respective incentives, roles and responsibilities, all coordinated by an independent national statistical office (NSO) and supported by a high-level government appointee focused on data – 'Chief Data Officer' (CDO).

Multi-stakeholder data partnerships have been proposed by a wide number of groups and are supported by new initiatives like the Global Partnership for Sustainable Development Data (GPSDD), but in many countries they are not happening. Concerns about methodological rigor, sample sizes and interoperability – as well as data security, privacy and ownership – prevent many governments from inviting nongovernmental actors into the national data production process. The risks are real and NSOs are right to be cautious, but the urgent need for more or better quality data requires a leap of faith. With the right, collaborative environments and robust oversight, there is no reason why common standards and protocols can not be set to guide the inputs of a wide range of different actors.

The report intends to support national and international policy makers taking an evidence-based approach to SDG planning and implementation. It looks to design modern, fit-for-purpose data systems that improve the quality, reach and impact of public service delivery and strengthen environmental protection. It sets out a vision for modern

data systems, supported by a broad range of actors, while also providing specific recommendations to drive urgent action.

There is no one right way to go about harnessing the data revolution for sustainable development, and there is not one perfect statistical system. The process of bringing on board a wide range of actors, each using different methodologies and approaches to produce, analyze, curate and disseminate data, will be messy and challenging. This report attempts to provide an independent view on how national systems should evolve in order to accommodate this change. But it requires an iterative approach, in which both national and international statistical systems are flexible and open to trialing new approaches.

The report has been prepared by an independent group of experts brought together by the SDSN. The members of SDSN TReNDS are all leaders in their respective data-related fields, drawn from across academia, government, national statistical systems, private companies, multilateral institutions and civil society. The group is chaired by three former members of the Secretary-General's Independent Expert Advisory Group on the Data Revolution (IEAG). Many of the members are institutional or individual members of the GPSDD, ensuring that recommendations build upon prior and emerging thinking on the data revolution for sustainable development from a broader community of actors.

ACRONYMS

CDO	Chief Data Officer
CEDA	Committee for Economic Development of Australia
CEPEI	Centro de Pensamiento Estratégico Internacional
CGD	Citizen-generated data
CIESIN	Center for International Earth Science Information Network
CIO	Chief Information Officer
CIMMYT	International Maize and Wheat Improvement Center
CoCB	Bogota Chamber of Commerce
COP	Conference of the Parties to the United Nations Framework Convention on Climate Change
CRVS	Civil registration and vital statistics
CSV	Comma-separated vales
CSO	Chief Statistical Officer
DALY	Disability-adjusted life year
DANE	National Administrative Department of Statistics (Colombia)
DDI	Data Documentation Initiative
DHS	Demographic and Health Surveys (United States)
DSG	UN Deputy Secretary General
ECOSOC	United Nations Economic and Social Commission
EO	Earth observation
GBD	Global Burden of Disease
GCI	GEOSS Common Infrastructure
GDP	Gross domestic product
GEO	Group on Earth Observations
GEO DAB	GEO Discovery and Access Broker
GEOSS	Global Earth Observation System of Systems
GIS	Geographic information systems
GPSDD	Global Partnership on Sustainable Development Data
HDX	Humanitarian Data Exchange
HLG	High Level Group for Partnership, Coordination and Capacity Building for Statistics
HOT	Humanitarian OpenStreetMap Team
HXL	Humanitarian eXchange Language
IAEG-SDG	Inter-Agency and Expert Group on Sustainable Development Goal Indicators
IB-NET	International Benchmarking Network for Water and Sanitation Utilities
IBGE	Brazilian Institute of Geography and Statistics
ICRAF	World Agroforestry Centre
ICSU	International Council for Science
ICT	Information communication technology
IDA	International Development Assistance
IEAG	Independent Expert Advisory Group on the Data Revolution
IMF	International Monetary Fund
IMIS	Integrated Management Information System

INEGI	National Institute of Geography and Statistics (Mexico)
IRDR	Integrated Research on Disaster Risks
IUSSP	International Union for Scientific Study of Population
JSON	JavaScript Object Notation
JUDS	Joined-up Data Standards initiative
LICs	Low-income countries
LMICS	Low- and middle-income countries
LSMS	Living Standards Measurement Study
MDGs	Millennium Development Goals
MICS	Multiple Indicator Cluster Surveys
MOOCs	Massive open online courses
NSDS	National strategy for the development of statistics
NSO	National statistical office
NSS	National statistical system
ODA	Official development assistance
ODE	Open Data Enterprise
ODW	Open Data Watch
OECD	Organization for Economic Cooperation and Development
OPAL	Open Algorithms Project
OPHI	Oxford Poverty and Human Development Initiative
PARIS21	Partnership in Statistics for Development in the 21st Century
PDAs	Personal Data Assistants
PES	Post-enumeration survey
RDF	Resource Description Framework
SDDF	State of Development Data Funding
SDGs	Sustainable Development Goals
SDSN	Sustainable Development Solutions Network
SDSN TReNDS	SDSN Thematic Research Network on Data and Statistics
UNGA	United Nations General Assembly
U.K.	United Kingdom
U.S.A. or U.S.	United States of America
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations Children's Fund
UNISDR	United Nations Office for Disaster Risk Reduction
UNSD	United Nations Statistical Division
WEF	World Economic Forum
WHO	World Health Organization
XLSX	Microsoft Excel Open XML Format Spreadsheet
XML	eXtensible Markup Language

1. Introduction: Data to help achieve sustainable development

In September 2015, 193 heads of state and government agreed to set the world on a path to a more sustainable future by adopting the Sustainable Development Goals (SDGs) via the 2030 Agenda for Sustainable Development ("2030 Agenda"). These goals are extraordinary; they represent the greatest consensus of the global international community since the signing of the Universal Declaration of Human Rights. But they are also ambitious. To meet global emissions targets, we need rapid transitions to green energy and nuclear power. To eradicate hunger and malnutrition, we need rapid expansion of nutrition-sensitive agriculture and fundamental changes in our consumer practices. To eradicate poverty for all, we need universal access to social safety nets, healthcare, employment opportunities and effective monitoring systems to ensure the most vulnerable are not being left behind. Each goal will require considerable investment by each and every government. Each will require strong political will. And each will require new technologies, data and innovation.

This report is about the data piece of this puzzle — in other words, the 'data revolution for sustainable development.' To achieve the lofty ambitions laid out by the SDGs — and the associated U.N. landmark agreements of the UN Framework Convention on Climate Change's 2015 consensus on climate change ("Paris Agreement") and Sendai Framework for Disaster Risk Reduction 2015-2030 ("Sendai Framework") — we need data and information systems that tell us what is happening in real time or in a timely way; enable us to understand people's unique vulnerabilities and challenges; enable us to see how services are working and whether they are reaching those most in need; and enable us to anticipate future opportunities, shocks, risks, and trends so we can adapt accordingly. Location is also a critical element, since location data and information provide insights for decision making now and in the future. Governments and citizens alike must be able to access this rich world of data in order to plan, organize and achieve their objectives; to hold each other to account; and to catalyze change while also ensuring personal liberty, security and equality of access.

Current data systems do not fulfill these ends. Poverty and basic health data, such as that relating to child stunting, is often five or more years out of date, while birth registration is often even older.[3] Administrative data like what children are learning, whether hospitals have enough medicine and whether people have access to transport are grossly underfunded in many parts of the world — if funded at all.[4] Worse still: Even when there is data available from third parties like civil society or private companies, it is often not used due to legal and institutional barriers, pre-conceptions about statistical methods and production processes, perceived quality issues or a lack of trained statisticians able to reconcile this data with official statistical records. But even if all this data were available, it would still be insufficient because it looks backwards. To respond to the immense problems of the 21st century, we should not only look to the past to learn lessons, but also look to the future to preempt challenges.

A 21st–century data system that is fit for purpose to both monitor and achieve the SDGs and the other U.N. landmark agreements should help governments to:

1. Plan and prepare for the future by anticipating climate change, environmental shocks and stresses, population dynamics, social challenges and changes, as well as new phenomena like mass urbanization and resilience challenges;

2. Manage and govern more effectively, providing policy makers with real-time or near-time information on the quality of services, the welfare of the population and the state of the environment so they can course-correct and change policies to meet changing demands;

3. Monitor historical progress and ensure we stay on course to meet our objectives, tracking changes over time and helping us to project where we are headed in the future.

Delivering these results requires a new approach to the development and management of data and information systems, which places data right at the heart of government.[5] Data needs to be the bedrock on which governments plan, budget, design implementation strategies, and improve their performance. Only with a data-based and evidence-led approach to decision-making can we have any chance of meeting the wide-ranging and very ambitious SDGs by 2030.

But placing data at the heart of government does not mean only governments can and should produce and/or share data. Private companies, universities, civil society and other third-party actors will need to contribute given the scale of the challenge. These partners can offer new skills, technologies, sources of data and analytical tools to improve our knowledge and understanding of sustainable development. Innovative (often privately-owned) sources of data can also provide a useful "check and balance" on government reporting, ensuring governments are fulfilling their commitments to the SDGs and are carefully managing the data at their disposal. Governments will continue to be central to the production of statistics, but as the range of data producers expands, governments' roles should morph from producer to coordinator of a broad data ecosystem. Statistical offices will transform themselves from information producers to knowledge builders. For example, they should be responsible for identifying useful, nongovernmental data sources that can help institutions and companies track the SDGs, and design policies and plans to achieve them. They should also assess the quality and reliability of third-party data and work to harmonize the data so it is broadly comparable with, or complementary, to official statistics, which will require capacity development.

That is not to say that partnerships are a silver bullet. Inviting more actors into the statistical production process and using new sources of data will create methodological challenges relating to sample sizes, differing methods and data interoperability, as well as raising important questions relating to data privacy, ownership and use. Poorly managed data partnerships risk exposing individual microdata (highly personal, individual data) to third parties who may not have the same developmental objectives in mind, or follow the same ethical principles official statisticians follow. These risks will have to be carefully managed by national statistical offices, as well as the executive branches of government. This is why one of several proposals of this report is that a new position within governments — the 'Chief Data Officer,' or CDO — be established to play a vital brokerage function: carefully sifting through alternative methods and sources of data, identifying quality partners and establishing partnership agreements with clear rules and expectations.

Notwithstanding the efforts made by the international statistical system since the adoption of the 2030 Agenda, the limitations of current national statistical systems for monitoring and achieving sustainable development are still relevant and well known.[6] This report attempts to suggest solutions for building more effective and efficient data ecosystems at local, national and international levels. It explains the kinds of data needed to achieve the SDGs and identifies the roles and responsibilities of different actors, as well as the urgent changes needed to build architectures capable of responding to the increasing demand for high-quality, disaggregated and geo-referenced data. Table 1 lays out a Theory of Change, summarizing the key actions required.

TABLE 1: THEORY OF CHANGE

	Priorities	Context	What must we do to bring about change?	Short-term changes	Long-term changes and beneficial impacts	Assumptions and risks
1	**Maintaining High-Level Commitment for Data**	In 2015, the 2030 Agenda for Sustainable Development ("2030 Agenda") was agreed upon, including 17 Sustainable Development Goals (SDGs). The agenda calls for more disaggregated data and data of a greater resolution to support the commitment of leaving no one behind.	Encourage governments to place data at the heart of decision-making. Create platforms for high-level, international discussion of the data imperative and the agenda to leave no one behind.	Improvements to data-informed planning and decision-making. Demonstration of value and impact of data (evidence to show investment in data is a smart move).	More effective policy design, more efficient use of resources and better outcomes. The data and evidence agenda in countries is a valued line of business with clear commitments, accountabilities and senior-level positions such as Chief Data Officers and Chief Statisticians, with an independent NSO in charge, etc.	Governments and global development initiatives see the benefits of data-based policy and decision-making, and are not more influenced by political dynamics. Governments are a trusted convener and collator of data and will not misuse or distort data for political ends. Governments appoint champions to lead and deliver on the process, notably appointing CDO, etc.
2	**Closing Persistent Data Gaps and Improving Data Quality**	There are acute data gaps affecting every country in the world (e.g. we have very poor knowledge of gender experiences of poverty). What data is available is often three or more years out of date; alternatively, the data that is available is patchy or irregular.	In partnership with organizations like the Partnership in Statistics for Development in the 21st Century (PARIS21), advocate for increased investment in statistical systems (including support for human capacity development) and show the return on investment of data. Show the value of innovative interim measures to help fill gaps and stress-test official statistics.	Increased awareness about the limits to our knowledge and the negative effects upon effective policy and management. There is more evidence on quality issues, data gaps, costs to fill them, and innovative methods to fill gaps or model data. Data is better used, and there is increased adoption of open data principles.	Well-trained, well-equipped statistical systems are established in all countries worldwide, generating more data of higher quality with greater frequency.	National governments, bilateral and multi-lateral donors have additional resources to make available for data system development. Money and training can resolve the current shortcomings in statistical systems. The U.N. and intergovernmental technical data processes are successful in setting standards that are implementable by countries for new data.
3	**Enabling Collective Efforts Among Data Communities**	Monitoring the sustainable development agenda is complex. No one entity can do it on its own.	Create spaces to engage a range of data producers and analysts from the private sector, academia and Chief Statisticians in the SDG challenge. Establish principles and standards to ensure data quality and reliability across diverse data producers, and to ensure data privacy.	Chief Data Officers are appointed to help governments identify new data sources and opportunities. The national statistical office's (NSO's) role evolves from producer of data to part producer and part quality controller over new data sources available. The UN Statistical Commission invites nongovernmental data producers as active participants.	NSOs work with CDOs, private companies, citizen groups, NGOs and academia to generate data through a more collaborative model that does not impose a strict divide between 'official' and 'non-official' data sources. Open data, data privacy and data interoperability are norms rather than exceptions.	All partners will respect data sharing and storage best practices, and will uphold data privacy rules. Third parties will produce reliable data over the duration of the SDGs that governments can count on and openly access.

TABLE 1: THEORY OF CHANGE (CONT)

	Priorities	Context	What must we do to bring about change?	Short-term changes	Long-term changes and beneficial impacts	Assumptions and risks
4	**Harnessing the Data Revolution**	We are in the midst of a 'data revolution,' brought about by technological change and the expansion of the internet (both geographically and in terms of content). This brings with it new data collection methodologies and approaches – potentially lowering costs of data collection tools like censuses, improving access and use of data, and enabling low capacity countries to leapfrog in adopting better data and data systems.	Pilot and make available at-scale new approaches to data collection, storage, management and sharing, drawing upon new technologies and private innovations. Build the capacity of governments and statisticians to use these new methods and to integrate them into the production of official statistics.	Common standards and replicable methods are established for the use of innovative new approaches to collecting, for example, population data or conducting infrastructure mapping, and use of geospatial data and Earth observations.	Countries are able to access a broad array of innovative interim approaches to data collection to help fill gaps in official statistics (e.g. using satellite imagery, radar or citizen-generated data), and have increased capacity to utilize and analyze these new approaches.	Countries and communities will welcome new methods and tech-based solutions to their data gaps. Governments are able to raise the necessary resources to build capacity and integrate tech-based solutions to their data gaps. Academia, civil society and private companies will work together to create standards and identify replicable best practices.
5	**Closing the Digital Divide**	There is a growing digital divide between those with access to the internet and new data technologies and those without, both across and within countries.	Support governments with money, technology, and training to: create an enabling environment for the expansion of broadband coverage, increase access to modern technologies and promote data literacy through data-oriented curricula in schools.	More people have access to smartphones, computers and high-speed internet. Governments create e-mechanisms to enable people to report on their personal circumstances or that of their environment using digital tools. More young people are being taught the potential of data, how it can be collected, how it can be used and its associated risks.	Future generations are empowered to understand, use, and protect data. All citizens in every country are invited to provide feedback on their services and governance using phones and e-based approaches (among other approaches).	Widespread basic data literacy will help promote its use for sustainable development policy and decision-making - and will help ensure responsive, safe data management.

2. The data we need: Understanding the past, present and future

A. MONITORING THE PAST

Tracking our progress on sustainable development requires huge quantities of data. In March 2016, the UN Statistical Commission approved a set of more than 240 indicators that they propose be used to monitor national and global progress towards the SDGs by 2030.[7] These indicators will come from a wide array range of data sources and tools including censuses, household surveys, civil registration and vital statistics systems (CRVS), administrative data systems, and environmental data such as geospatial imagery.

Few countries in the world have statistical systems ready and equipped to monitor the breadth of the sustainable development challenge, as an analysis of population data serves to show – see Box 1. As this report will explain, investments in statistical systems and new data partnerships and innovative methods will be essential, but we must also attempt to use what resources we have more effectively. Nearly every country in the world has an existing statistical system that conducts the national census, compiles household surveys or prepares national accounts, even if infrequently. Each of these statistical processes generate data that – even if two, three, five or 10 years out of date – can give us insights into a country's past and present. However, in a great many cases this historical data is poorly stored, relegated to a report gathering dust on a shelf. In addition, data is often collected in a piecemeal fashion with little attempt to assimilate complementary data and identify trends across data sources.

But to achieve the SDGs and be able to monitor our progress over time a historical perspective is essential; countries need to be able to compile a baseline (see Box 1), need to be able to analyze trends over time, and to predict future trajectories. Legacy data and data systems are frequently ignored as people look to the data revolution and the new, shiny tech-based approaches. But with relatively modest investments in digitizing, cleaning, and standardizing data, it is possible to derive value from these historical data. The process of digitizing, cleaning and standardizing data is also important so that new data can be consistent with that compiled in the past, enabling us to track progress over time and to assess our trajectories.

A case study of the measurement of drinking water supply in rural Bangladesh serves to demonstrate the value that can be derived from utilizing legacy data systems (Box 2).

B. MANAGING THE PRESENT

Fundamentally, data for sustainable development should help governments to effectively manage their resources, services and responsibilities so as to provide the best possible support and protection for their citizens and the natural habitat. Data should serve as an administrative tool supporting governments to make judicious decisions about where and how to direct attention and resources. But to be a helpful tool, data needs to be both timely and relevant. Data that is three or more years out of date cannot help a government make an effective decision between investing in one health clinic over another (though historical data can help governments identify where there have been services over time). Conversely, having access to real or near-time data can help governments make nimble decisions about how to move capacity and resources with the potential for huge efficiency gains. This kind of data in support of management and administration is known as administrative data. It is data that tells us how things tick and helps us to run effective and responsive operations, services and businesses.

In a wide array of literature on the data revolution and in national strategies for the development of statistics (NSDSs), administrative data is highlighted as the single greatest area of systematic underinvestment.[16] The returns on investment are immense, but few cash-strapped, low-income countries are inclined to invest in long-term systems-building when there are other competing, urgent needs. Furthermore, few donors are inclined to invest in administrative data collection methods and tools as these systems are managerial and process-oriented, not

responding to a specific problem or providing an immediate tangible solution. This makes it harder to explain the returns on investment, both socially and economically.

Building robust administrative data systems in this resource-constrained environment is therefore dependent upon two things:

1. Making available data interoperable, and

2. Using diverse sources of data or 'multi-modal' data collection methods.

Data interoperability is the ability to convert or store data in a format that is easy to use and distribute, so as to facilitate the easy exchange of data. Within governments, interoperable data is essential for data to be shared across line ministries and departments in comparable, useful formats. This enables integrated program design and monitoring that cuts across sectors. Particular important is mapping data, which if made available in accessible formats by mapping agencies or ministries of land or environment, can be used by the NSO to assess the spatial distribution of poverty or wellbeing indicators.

Interoperable data is particularly important when looking to provide services to vulnerable groups. Take, for example, the situation of a vulnerable child. For this child to be recognized by the government and for their welfare to be tracked over time, they first and foremost need to have their birth registered for a record of their identity. Ministries of health most commonly collect this data. They then need to go to school, and the ministry of education should know if this highly vulnerable individual is able to access public schooling. There should be a record of their care situation (whether living in an institution or with a foster family) and an address. And there should be records of their developmental progress, health and wellbeing. All of this information is collected by different sections of government, often with the support of third parties such as UNICEF or the World Health Organization (WHO). If the data is not recorded in systematic administrative data systems within ministries and government departments, it will be nigh on impossible to monitor and track the welfare of that child over time. Furthermore, to ensure the child receives a holistic program of care, that data needs to be shared across departments – requiring that it be interoperable.

Creating truly integrated and interoperable data systems with real-time data sharing across governments relies on records held electronically; front-line service agents with access to computers and the internet (or the capacity for the department or ministry to digitize records); and agreements across government on data exchange, standards and storage. For third-party data to be integrated into this data architecture, they also need to agree on standards for data collection, storage and ease of use. These processes are not in and of themselves complex, but they are time- and resource-intensive, requiring high-level political commitment to bring about systemic change.

Given the gaps in government administrative data systems, multi-modal data collection can be a useful resource. Multi-modal data is when two or more data sources are overlaid with one another to offer a more complex picture of a community or geography than might be provided by any one of them, helping fill gaps in other sources. Examples include satellite imagery overlaid with telecommunications data to map population movement, or satellite imagery overlaid with citizen-generated data to carefully map facilities or risks within a given community. For more on the value of multi-modal data collection, see Box 3.

"We are determined to protect the planet from degradation, including through sustainable consumption and production, sustainably managing its natural resources and taking urgent action on climate change, so that it can support the needs of the present and future generations.

Transforming Our World: The 2030 Agenda for Sustainable Development[18]

C. PLANNING FOR THE FUTURE

A key imperative of sustainable development is to protect the Earth's natural resources for future generations, ensuring that we do not deplete natural stocks at a rate that cannot be replenished. Sadly, this has not happened throughout recent history. With a world population now at 7.2 billion people and an annual gross domestic product (GDP) of nearly USD 90 trillion, the world economy using today's technologies is already exceeding several of the Earth's "planetary boundaries."[19] Without access to sexual and reproductive health services and other targeted responses, the global population will rise to 9 billion people — or possibly more — by 2050, and to 10 billion before 2100. [20] Many natural resources and ecosystems essential for human and societal wellbeing are already under threat, and will be further threatened or destroyed if current generations do not consume them sustainably. The world will experience unprecedented crises of food production, public health and natural disaster, among other threats. Food prices will soar, and some parts of the world may be rendered virtually uninhabitable as a result of climate change and water stress.

Managing these risks and the increased incidence of natural disasters arising from extreme weather requires that we use data to analyze past trends and predict future scenarios. The necessity to effectively plan for and manage risk is well articulated in the Sendai Framework, and must be pursued in concert with the SDGs (see Box 4).

Sendai identified the need for new data capacities that enable governments to project into the future, anticipating our trajectory and changing course as required. The Sendai Framework specifically calls for enhanced scientific work in disaster risk reduction and a better coordination of existing networks and scientific research institutions, enabling more fluid exchange of data, modeled assessments and policy recommendations.[21]

A key tool for this kind of analysis is forecasting or modeling future scenarios. One such example is The World in 2050, a project assembling leading modeling teams to perform an integrated assessment addressing the full spectrum of sustainable development challenges. The key value of the project is that it maps our trajectories on a range of issues or goal areas and looks at the synergies and tradeoffs among these issues — for example, identifying how rapid electrification might result in excessive non-renewable energy use. This kind of approach allows governments not only to predict future trends, but also to make informed policy decisions that take into account potential sacrifices.

Unfortunately, few countries in the world have this kind of technical capacity within their statistical systems. However, academia and private industries like insurance often specialize in such methods, and can be a powerful partner for governments if they are invited into the data collection, policy development and planning processes. This relies upon governments having a more open and responsive attitude to data partnerships, in which scientific predictions and forecasts are given equal weight to current and past analysis within the data-based policy and decision-making processes. International entities like the GPSDD and SDSN can also play a useful role, showcasing examples of successful public and nongovernmental collaborations to help forecast scenarios and design responsive policies and programs.

BOX 1: THE CHALLENGES OF ESTABLISHING A POPULATION BASELINE TO MONITOR THE SDGS

The adoption of the SDGs by the global community in 2015 represents a concerted attempt to ensure sustainable and equitable development over the period from 2015 to 2030. Meeting the goals and their associated targets will be challenging, particularly in the context of the overarching principle that no one should be left behind. However, there is a real risk that the efforts made to meet the goals and targets may be compromised by the inadequacy of the data, particularly in the Global South, to benchmark, monitor and track the 232 unique indicators[8] of progress in meeting the SDGs.

Of the 232 indicators, 86 (37 percent) require population data in both numerator and denominator, while a further 11 require population data in one or the other. Consideration of those 97 indicators suggests that 92 of them will require data of the kind currently collected in censuses (and the population projections derived therefrom) and nationally-representative surveys. Data from civil and vital registration systems will be required to monitor 15 indicators. With each source of data, significant barriers will have to be overcome.

In order to understand these barriers and challenges, we need to consider for each of the major potential sources of data their usefulness for providing baseline data, as well as for continued monitoring.

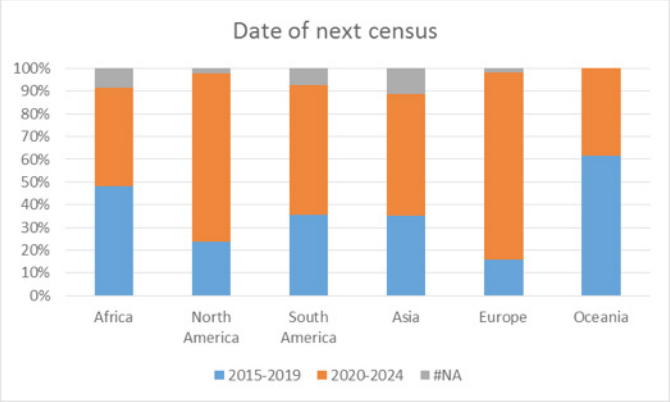
CONCERNS ABOUT DATA SOURCES

Census data and population projections

For most countries in the Global South, the national census and the population projections they are based on will be central in providing both the baseline and the monitoring data of population counts, especially at the fine levels of disaggregation required to ensure compliance with the principle that no one be left behind.

Only a very few countries enumerate their populations more frequently than every 10 years (and those that do so are typically small island states or the developed countries of Europe, the latter of which derive their population estimates from administrative systems). The long lead time to include and test new questions means that it is unlikely that any census to be conducted in the first five years of the SDG period will include new questions that will allow the benchmarking of SDG indicators, especially at fine levels of granularity. According to the UN Statistics Division, nearly half the countries in Africa are expected to conduct their 2020 census between 2015 and 2019. For those countries that will conduct their 2020-round census between 2020 and 2024, even if new questions are included the data will at best be applicable to the middle third of the SDG time period.

The countries that will conduct their 2020-round of censuses in the period 2015-2019 will therefore likely only be able to include new questions in their censuses conducted from 2025 to 2029, at the very end of the SDG time period. Furthermore, for countries that conduct decennial censuses, approximately half will run only one census between 2015 and 2030. Taken together, this means that earlier data or the results of prior population projections will have to be used to establish the baseline data to monitor the SDG indicators.



A second substantive concern relates to the quality and granularity of the data collected in a census. Censuses, particularly in the Global South, are prone to under-enumeration, the effects of which are sometimes compensated for by means of a post-enumeration survey (PES) to assess the extent of the undercount. The PES also allows for the results to be scaled to reflect the true population size and characteristics. The Statistics Division does not strongly recommend PESs due to their technical and logistical complexity.[9] But in the context of the "leave no one behind" principle, an accurate enumeration of the population at fine levels of disaggregation becomes more important. Despite their intrinsic complexity, PESs (or alternative methods of correcting censuses for under-enumeration) should therefore be more strongly advocated.

While other methods of data collection are being explored and developed (such as remote sensing or integrating data from a number of heterogeneous sources) and hold great promise, these are still in their infancy. Many of the indicators will continue to rely on data collected in censuses, vital registration systems or surveys sampled off of a census sampling frame.

Nationally-representative surveys

Nationally-representative surveys, whether part of an international program such as the Demographic and Health Surveys (DHS), the World Bank-run Living Standards Measurement Study (LSMS, which provide data, inter alia, on poverty and socio-economic wellbeing) or the Multiple Indicator Cluster Surveys (MICS) run under the auspices of UNICEF; or those conducted in country (for example, labor force, or general household surveys) will have to be used to provide the data for a great number of indicators, as the detail of the responses required (for example, on Gender-Based Violence) are not suitable for collection in censuses. In aggregate, surveys of one form or another might be expected to provide data for the baselining and continued monitoring of 65 indicators.

Several important issues flow from this observation. The first is that harmonized and standardized questions on many of the indicator topics have not been agreed, and many of the topics required by the indicators may not be routinely collected in the surveys that have been run in the past. Existing surveys will have to be expanded to include new questions, or entirely new surveys will have to be designed, piloted and run.

Second, the census (or population projections derived from censuses) provides the essential sampling frame for drawing nationally-representative samples. Census data or population projections will have to be able to provide accurate sampling frames down to very fine levels of disaggregation to allow accurate sampling of small populations. This again compounds the need for a program of PES or for a concerted program to develop alternative methods and approaches for estimating population counts (by age, sex, gender, income, race, ethnicity, disability, etc.)

Third, in order to achieve the granularity desired by the principle of "leaving no one behind," the size of the surveys will have to be increased dramatically (with due consideration of quality and sustainability) to produce estimates of sufficient precision with a commensurate increase in cost, although some of these costs may be ameliorated through the adoption of new data collection technologies. A further cost will arise from the fact that the periodicity of many of the surveys will have to be increased to permit regular monitoring of progress in meeting the SDGs.

Civil and vital registration systems

Throughout the Global South, civil and vital registration systems (CVRSs) remain incomplete, often substantially so. The Statistics Division's Population and Vital Statistics Report provides information on the completeness and timeliness of data on reported births and deaths. At the beginning of 2017, no country in continental sub-Saharan Africa had birth registration that is estimated to be more than 90% complete, and for a significant number of countries, the data that have been provided are more than five years out of date.[10] For many of these countries, estimates of the number of births are themselves derived from population projections rather than from civil registration systems. The completeness of reporting of deaths is generally regarded as worse than that of birth reporting, and many countries have not been able to report numbers of infant deaths at all.[11]

Other regions have generally better CRVS data, but the challenge of improving the quality of CRVS data pertains across the Global South. While the need to improve the quality and timeliness of CRVS data has been acknowledged for some time and steps are already being taken to do so, it may still be several years before the fruits of these initiatives are visible.

Source: Written by Tom Moultrie, International Union for the Scientific Study of Population (IUSSP)

BOX 2: DESIGNING INTEROPERABLE DATA SYSTEMS TO REDUCE RISK IN RURAL DRINKING WATER ACCESS AND SUPPLY IN BANGLADESH

SDG Target 6.1[12] provides an expanded framework to capture the multiple and intertwined factors of water quality, supply, reliability, affordability, equity of access and use and infrastructure functionality. Measuring these components effectively is costly for systems that have multiple decision needs and levels of governance, especially federal systems where monitoring responsibilities are assigned by state boundaries. Therefore, nationally-aggregated databases are rare or, when in place, are often in silos by category such as water quality parameters, rates of access and consumption, or infrastructure inventories. However, many of these data points often already exist, but are spread across water utilities, municipal governments and/or water regulators.

Production of administrative data occurs across institutional levels, and is increasingly seen as a meaningful resource to be reconciled and aggregated. Doing this in a cost-effective and standardized manner is increasingly feasible. This approach is already undertaken manually by the Government of India's Integrated Management Information System (IMIS), which systematically integrates household level data on population, water scheme type, water sources and water quality with disaggregated administrative data on financial planning, budgetary allocation and monthly progress reports.[13] In its 2017 guidance note and progress report, the WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene identified programs underway to reconcile administrative data for the SDG monitoring framework from countries with highly regulated and standardized water sectors.[14] This is primarily from the urban water sector and builds on previous successes from the World Bank's International Benchmarking Network for Water and Sanitation Utilities (IB-NET).[15]

Recent work by the Sustainable Development Solutions Network's Thematic Research Network on Data and Statistics (SDSN TReNDS) has identified the potential to apply this approach in Bangladesh, a country that faces a unique convergence of water-related risks. Data points from the government's sub-national monitoring processes, under the supervision of under and district committees and government engineers, provide an opportunity to monitor and report the provision and safety of public water infrastructure using existing data points. If reconciled with household surveys and the network of government laboratories testing water quality, this could be a first step towards a strengthened water safety monitoring network and one that would help triangulate national official statistics. The politics of access, potential inconsistencies of data structure and complexities of local government reporting remain key hurdles. But lower transaction costs and faster, less time-intensive approaches are opening the doors of possibility.

Source: Written by Alex Fischer, PhD candidate, University of Oxford

BOX 3: THE PORTFOLIO APPROACH TO INVESTING IN DATA SYSTEMS

Before the data revolution, governance data systems tended to be designed in a way that applied the single best approach to measurement for a specific data need. When data collection was comparatively expensive and technology change comparatively slow, such an approach served well. The methodology for counting population through a national census has remained robust and powerful for most of the last millennium. Today, however, change is fast and costs are falling (though unevenly). Moreover, much more is expected from data systems; instead of serving a single governance function, they more often now are required to support many policy communities. Therefore, data systems that generate the most value for the investment should be designed in a manner that aims at a package of data collection strategies, producing the best value-for-cost combination. And when such a strategy is pursued, it may be possible to reap an additional benefit. In much the same way that modern portfolio theory has shown how investors can improve risk-adjusted returns by combining distinctive asset categories, architects of data systems are discovering that they get better results by combining multiple approaches to data collection.

The implications are profound. Those charged with designing data systems in the past had to know one thing well; today, they must be comfortable with a broad array of technology and institutional choices. In the past, there was a premium on consistency and stability; today the imperative is to adapt to change fast enough to take advantage of innovation, but without triggering unhelpful disruption and distrust. As a result, the task of designing a fit-for-purpose data system is increasingly one that only a purposive community can take on, because no single individual or organization can plausibly have the right information.

In the private sector, most leading firms have responded to the rapidly changing data technology landscape by concentrating responsibility for strategic planning with respect to information systems in the position of a Chief Information Officer (CIO), as opposed to the head of Information Technology department. Within firms, CIOs are able to map organizational needs to the broad landscape of information technology in order to design data and analytic systems that add value. Where such strategies thrive, they can go even further, as in the case of the innovations in information technology and informatics developed by Jack Levis at UPS which have been so successful that the core work is baked into the entire DNA of the firm (see for example UPS' ORION project)[17].

Few countries or international organizations yet have CIOs, but all need to start building the kinds of capabilities associated with them. Successful data decisions today require carefully calibrating a portfolio of measurement solutions to meet decision-making needs.

Example: High-resolution population mapping in northern Nigeria

When a Bill and Melinda Gates Foundation initiative to eradicate polio began working in northern Nigeria, it quickly became apparent that existing data on settlement and population distribution was not appropriately suited to the decisions that had to be made. There was inconsistency, incompleteness and inaccuracy regarding the location and names of settlements. The population counts in each settlement were not up to date and, in some cases, highly inaccurate. As a result, it was impossible to design and implement an effective vaccination campaign.

No single approach to measurement would have been capable of meeting decision support needs. Field surveys were too slow and prone to error. High-resolution imagery could detect structures but not generate estimates of populations. By combining these two sources of information, along with other data such as road networks and land cover maps, it was possible to generate highly accurate estimates of population distribution in less than three months. These new estimates were instrumental in the effort to deliver vaccines where and when they were needed.

Image 1: High-Resolution Imagery Helped Create More Accurate Population Estimates in Northern Nigeria

Source: Written by Marc Levy, Center for International Earth Science Information Network (CIESIN)

BOX 4: A SYSTEMATIC APPROACH TO MANAGING FUTURE RISK, USING SCIENTIFIC METHODS AND DATA

The recent synchronous adoption of landmark UN agreements the Sendai Framework for Disaster Risk Reduction, Sustainable Development Goals (SDGs), COP21's Paris Climate Conference and the World Humanitarian Summit and Habitat III (focused on housing and sustainable urban development) has created a rare but significant opportunity to build coherence across different but overlapping policy areas. This coherence will serve to strengthen existing risk fragility and resilience frameworks for multi-hazard assessments, and aim to develop a dynamic, local, preventive, and adaptive urban governance system at the global, national, and local levels. For example, taken together these frameworks make for a more complete resilience agenda as building resilience requires action spanning development, humanitarian, climate and disaster risk reduction areas.

The agreements represent a major turning point in the global efforts to tackle existing and future challenges in all countries. Specific emphasis is apparent to support resilience-building measures, and a shift away from managing crises to proactively reducing their risks. In order to respond efficiently to all of the agreements, synergies between policies, programmes and institutions, need to be highlighted and supported by the alignment of actions. This must include integrated data systems and forecasting. With the collection, analysis and monitoring of data as a requirement of all the landmark agreements (with varying levels of complexity), there is an opportunity for policy makers and practitioners alike to ensure minimal duplication and greatest impact of the data collected. Scientific methods, stakeholder networks and communication all offer critical assistance to the development of well-informed policies and decisions across all countries and stronger linkages between evidence and decision-making in policy and planning are also needed to ensure delivery of the 2030 agenda for sustainable development.

Key recommendations in building coherence between these agreements and agendas include:

- Raising awareness and understanding with national and sub-national governments on how the different frameworks align is critical; the relative political weight of frameworks may affect collaboration and coherence.
- Facilitating key partnerships which help avoid duplication and maximize gains. Institutional incentives to work together may also be required to reinforce joint working across agreements.
- Instituting clear governance arrangements to ensure successful collective action and accountability.
- Developing consistent definitions, particularly on resilience and risk, that connect as common themes across all of the agreements.
- Promoting science and technology involvement by funding national/ regional research projects. The Sendai framework specifically calls for enhanced scientific work in disaster risk reduction and a better coordination of existing networks and scientific research institutions.
- Joined up monitoring processes which track progress on implementation of the frameworks. This will also help minimize the reporting burden on countries, making data collection achievable.
- Ensuring national ownership and leadership on all of these frameworks will also be fundamental to success.

Source: Murray, V. et al. (2017)[22]

3. Actors and incentives in the current data landscape

There is a wide range of expertise required for comprehensive SDG planning, management and monitoring. Classically trained statisticians may be experts at conducting surveys or censuses, but they are less likely to have expertise in administrative data collection and analysis within a given sector – such as interpreting satellite imagery or geographic information systems (GIS) – or forecasting. The range of expertise required can sometimes be gathered from across government, with the NSO undertaking the census and surveys, ministries of environment or agriculture interpreting satellite data and using GIS, and different sectoral ministries compiling administrative data. Still, more often than not (and particularly in countries with low government capacity), these kinds of national statistics need to be gathered in partnership with nongovernmental actors who can provide additional expertise or resources.

A. ENGAGING A BROAD ARRAY OF ACTORS: LESSONS FROM HEALTH MONITORING

We can trace the many actors who can contribute to a more robust and actionable data system by examining a single health measure like mortality. In high-income countries or countries with robust civil registration and vital statistics systems, the preferred source of mortality data is measurement through continuous registration of deaths (administrative data). In such a system, administrative data are collected in hospitals, mortuaries and by professional groups such as health visitors, and are centrally compiled in a civil registration and vital statistics system (CRVS) that is managed by the ministry of health or other departments of central or subnational government.

But for many countries – particularly those with higher levels of mortality – administrative and centralized statistical collection capacity is weak. As a result, household surveys may serve as the principal vehicle for data collection for vital statistics, among many other demographic indicators. These surveys should be conducted by the government, but because of their high costs and the need for specific expertise to administer the survey, they are often supported by third parties such as international bodies or donor governments. Examples include the UNICEF's Multiple Indicator Cluster Survey (MICS) program, the World Bank's Living Standards Measurement Study (LSMS) program and the U.S.A.'s Demographic and Health Survey (DHS) program. While this system of household survey collection to determine vital statistics does not provide the same timely accuracy of information as a robust, real-time administrative system, partnerships create a level of functionality that can be built upon for further use. Furthermore, the significant in-country technical assistance provided to national statistics offices by these programs builds statistical capacity. This increases sustainability and country ownership, enabling countries to more quickly achieve data independence in the form of a robust and actionable national data system.

Once collected, data may be cleaned, further verified and given new utility by other expert nongovernment actors. Outside the bounds of individual government statistical systems, experts from academia or multi-lateral institutions can create tools and models based on existing data that go further than any one dataset alone. The Global Burden of Disease (GBD), supported by the WHO and other organizations, is an incredibly useful tool for secondary analysis of mortality data. For example, GBD uses models such as the disability-adjusted life year (DALY) to form resources for governments and multilaterals through which they can weigh evidence on diseases and risk factors and craft the priorities and investment decisions necessary for a tailored public health agenda. Expert-modeled data is crucial when trying to fill data gaps and to identify trends, although it cannot be a substitute for empirical observation and building local statistical capacity (see Annex 1).

In addition to governmental, supranational and expert actors, private companies and civil society organizations often have a role to play in data collection. Beyond the mortality example, we see that in some high-income countries, health data can be extracted from private insurance providers or from

internet search companies like Google that can use big data to identify patterns in people's health preferences, enquiries and ailments.

Although these kinds of big data are not an appropriate primary source in the case of mortality estimation, they can be useful to monitor other health-related SDG indicators such as communicable and non-communicable disease burden, air pollution and harmful use of alcohol. They may be useful for "nowcasting," or predicting current trends, such as when a bout of influenza is about to spread and its distribution patterns.[23] Big data analytics also have the potential to add another layer of comparable data or to triangulate existing official data, allowing us to better assess data quality and more easily disaggregate data.

Nongovernmental actors such as civil society groups and universities can also play a very important role monitoring sustainable development issues when the government either cannot or will not. For example, in many countries, conservative or religious factions in government may prevent the monitoring of safe access to family planning and abortions. SDG Target 3.7[24] guarantees universal access to sexual and reproductive health care services, including family planning, in all countries around the world. To fulfill this objective, nongovernmental actors may have to play a central role in monitoring family planning services and helping to deliver services to those most in need. It is also imperative that they have means by which to communicate this data to local civil society groups and to the international community – for example, through multi-stakeholder forums during the annual meeting of the UN Statistical Commission – so that governments can be held to account using this third-party evidence.

B. MAPPING ROLES AND RESPONSIBILITIES

As the range of sustainable development challenges facing governments widens and deepens (with mounting population growth, urbanization, accelerating climate change and so on), broad partnerships like these will be essential. Nongovernmental partners have the potential to: contribute deeper expertise in a given methodology or technology; provide large quantities of data that may enable deeper disaggregation or the measurement of under-measured issues; and provide data that can be used to cross-check the quality and accuracy of official statistics. Bringing them into the national process can help to ensure their expertise contributes to the official statistical production process, rather than creating duplicative measures or processes.

Table 2 provides a summary of the range of actors involved in the monitoring of sustainable development, including a general description of their expertise, potential roles and responsibilities, as well as their incentives.

C. ANTICIPATING SHARED RISKS

The examples in Table 2 highlight the beneficial role that nongovernmental actors can play in producing data that supports sustainable development. If carefully directed, governments and the global data community will be able to tap a deep pool of public and private data sources and expertise that can support SDG monitoring and the achievement of sustainable development. If not carefully directed, however, the 21st century proliferation of data, data producers and data users could lead to a data 'dystopia'. In this alternate future, official statistics will become a smaller and smaller piece of the data pie, with the majority of data being compiled by private companies and viewed as a monetary commodity. Only those with money would be able to access the data holdings collected by private companies, leaving governments and their national statistical offices to use outmoded tools to gather data that are less timely – and in some instances, less accurate – than that being gathered by nongovernmental actors.

This scenario may seem extreme. But consider a company like Facebook: At the end of 2016, it had 1.8 billion users.[25] That means approximately one-quarter of the world's population was using a social media platform that enables the parent company to: see their location, their friends and relationships, their age and their educational qualifications; use photos to assess their well-being; and more. Platforms like Facebook are becoming more and more commonplace, increasing the amount of socioeconomic data that is in private hands. We, as individuals, cannot see this data; have little to no control over what the private, parent company chooses to do with this data; are excluded from corporate-level decisions that might involve selling our agglomerated data to third parties, and so on.

The sustainable development agenda provides an opportunity to build national statistical capacity so

that NSOs have the resources and technical skills to monitor the broad suite of sustainable development issues. But it also presents an opportunity for NSOs and other actors within the national statistical system (NSS) to engage private actors; ask for their help monitoring such a broad, complex range of issues; welcome their technologies and approaches; and work together to establish a consensus on data transparency, openness and citizen privacy, making private data work towards the public good. It also provides an opportunity to test the scalability of many of the new approaches: to see if they are transferable across contexts, and whether they work not just for private companies or pilot studies, but also for whole-scale monitoring of large areas or population groups.

Sustainable development is a 21st-century imperative. To achieve a more sustainable, inclusive and resilient world, we will need a wide array of data that enables us to see how, when and where things are happening – and to whom. Given their breadth, resources, expertise and capacity for innovation, private companies, universities and civil society groups all have to be at the table. Governments should act now to engage with the ever-increasing range of data actors and to bring their skills to bear on the sustainable development challenge. Cross-sectoral and cross-disciplinary collaborations will not be easy — there are a broad range of institutional and political factors that may slow down these kinds of partnerships – but with clear common goals, strong leadership and creativity, hurdles can be overcome. The next chapter maps out common challenges and identifies creative responses to motivate action now.

TABLE 2: DATA PRODUCERS: CONTRIBUTIONS, INCENTIVES, ROLES AND RESPONSIBILITIES

Data producers	Expertise	Incentives	Roles	Responsibilities
National statistics offices (NSOs)	Data collection and compilation of official statistics from the census, surveys, or administrative systems. Knowledge of international standards and methods.	To ensure government policy is informed by quantitative and (in some cases) qualitative evidence. Receiving funding from central government, and funding from the international community.	Production of essential official statistics such as population and vital statistics; national accounts; and other social, economic, and environmental statistics. Coordination and expert review of other data from across government and from non-governmental actors.	To ensure high-quality, consistent production of data on government performance and data for SDG monitoring.
Other government departments	The compilation of sector-specific administrative data (such as health or education information) or technical skills using certain software like ministries of planning or environment using geographic information systems (GIS).	To develop data systems that can help improve service performance and efficiency and/or target resources most effectively. Receiving funding from central government and the international community.	Production (or co-production with the NSO) of sector-specific datasets relating to health, education, agriculture, etc.	To use data to improve service delivery and effective, responsive governance. To share these data with the NSO and across government.
United Nations agencies, the World Bank and the International Monetary Fund (IMF)	Technical expertise in survey methodologies and the collection of sector-specific statistics, including poverty and economic data. Detailed knowledge of international standards and cross-country approaches.	To improve the quality of internationally comparable data, to target resources most effectively. To show leading sectoral expertise and access central U.N. funding.	Technical advice to NSOs. Verification and standardization of national-level data into internationally comparable estimates. To facilitate agreements on standards, etc.	To support national processes and build local capacity, rather than parachuting in with technical skills.
Bilateral donors	Expertise derived from the production of official statistics within their own high- or middle-income countries. Targeted investments to improve program delivery or organizational efficiency. Data-based or informed program design and delivery.	To build systems capable of tracking investment impact. To ensure more responsive, accountable and transparent government partners. To create a culture of transparency and data-sharing to attract foreign investment and minimize risk.	To fund statistical capacity development. To identify new international data partners. To facilitate peer-to-peer learning between statisticians across countries (in the Global North and Global South).	To fund basic, core statistical capacities where the government is unwilling or unable. To support national processes and build local capacity, rather than parachuting in with technical skills. To provide resources to support the independence of the NSO.
Universities and academia	Sector-specific data production, data cleaning, and analysis. Highly trained and qualified specialists.	Funding and high-level recognition and use of their work.	Specialist data compilation and/or verification. Providing additional capacity where there is a government shortfall.	To interrogate official statistics against official standards and offer supplementary or complementary evidence.
Private companies	Market and consumption data, expertise in managing big data and big data analytics.	Philanthropy and/or public service. Public marketing and branding. Government contracts.	Providing consumer data and/or big data analytics to complement and/or supplement official statistics.	To make open and available non-sensitive social, environmental and economic data. To maintain respect for individual data privacy and to agree upon an ethical data-sharing framework.
Citizens and civil society groups	Community-level data collection (often qualitative), citizen-generated data, data mapping, data literacy and/or methodologies not being pursued by government. Data communication and data-based advocacy.	To provide a check on government and private company data collection, utilization and communication. Public recognition. Changing policy and mindsets.	To produce complementary evidence (often qualitative) to cross-check official statistics and make them more accessible to ordinary citizens. To produce a deeper layer of analysis that humanizes statistics.	To interrogate official statistics and offer supplementary or complementary evidence. To monitor issues the government is unwilling or unable to measure.

4: Achieving a modern data system: Breaking down institutional barriers and fostering new partnerships

There are a number of challenges to establishing modern data systems and broad data partnerships for monitoring and attainment of the SDGs. Challenges include differing incentives; capacity issues; a lack of formalized spaces for multi-stakeholder engagement; differing standards and rules about data sharing; and varying levels of resources. This section highlights four pathways to overcoming institutional roadblocks, presenting practical suggestions that aim to challenge the status quo. It also includes illustrative case studies to show groundbreaking initiatives already underway. The four pathways correspond to and build upon the areas of action identified in the IAEG-SDG's report "A World That Counts",[26] but they also relate to the five principles identified in the Theory of Change (Table 1). These are explained further in "Counting on the world: a roadmap for urgent action" (Chapter 5).

A. GOVERNANCE AND LEADERSHIP

To date, discussions on data and monitoring of the SDGs have taken place predominantly within national statistical offices and their global governing body, the UN Statistical Commission. But as highlighted in "A World That Counts" and expressed at various events during the 48th Session of the UN Statistical Commission, NSOs cannot work alone. Inadequate financing, limited technology, lack of human and institutional capacity, and ever-growing demands beset many. Partnerships with nongovernment actors are imperative to capitalize on new data sources, technologies, and approaches. The Cape Town Global Action Plan ("Global Action Plan") recognizes this, calling on national governments to revise statistical laws and regulatory frameworks to develop a mechanism for the use of data from alternative and innovative sources within official statistics.[27] This recommendation is important and deserves to be monitored at the highest level over time, either within the Statistical Commission or through the Voluntary National Review process whereby countries report on their progress implementing the SDGs at the annual High-level Political Forum. To maintain a strong spotlight on data-related issues and evidence-based policy-making, governments should also: look to broaden and raise the profile of the current national statistical system; work with the U.N. to create a standing high-level panel on data; and ensure there are forums for other data contributors to engage with and be held to account for their contributions to the monitoring and achievement of the SDGs.

BROADENING THE NSO AND RAISING ITS POLITICAL PROFILE

NSOs are very diverse institutions in terms of capacity, resources, their responsibilities, their leadership, and legal frameworks. Nonetheless, a number of entities – including the Organization for Economic Co-operation and Development (OECD) – have recommended that NSOs seek to encourage new data partnerships and the use of data from alternative sources by changing their core role and responsibilities. A common recommendation is that NSOs evolve from producers of data to coordinators of the broad data ecosystem, responsible for identifying a wide range of data sources and assessing their quality and rigor before using this data to compile national statistics. The OECD goes so far as to suggest NSOs become 'clearing houses' of data, responsible for certifying new data sets and methods.[28] The advantage of such an approach is that the NSO no longer has to compete with third-party actors, instead actively collaborating, nor does it need to house the latest technology or methodological approaches. The NSO will have an essential function regardless of how many data actors emerge, though they will need a strong mission and support from the executive (Head of State) to coordinate such a broad range of actors and ensure the data sources that they sanction are the primary sources of official statistics. Furthermore, NSOs will still need to retain a data production function, at least for the

foreseeable future – for example, overseeing the administration of the census and the production of many household surveys. Serving both functions will require additional investments in NSO capacity, as well as a retooling of current NSO staff to better equip them to serve a quality control function over the broad data community.

To cope with the exponential growth in data producers and the boom in potential data sources for monitoring sustainable development, countries should create a position of Chief Data Officer (CDO).[29] The CDO should work alongside chief statisticians, NSO staff and stakeholders to survey, catalogue, and harness the broad swathe of non-official data available from third parties and to propose ways in which new sources of data and information can be brought into the national SDG monitoring process. This might include identifying promising, high-quality citizen generated data, or identifying new, open sources of geospatial imagery which can be overlaid with official statistics (and fostering better collaboration with national mapping agencies, where they exist). Much as CDOs in private companies oversee all the bits and bytes of data that flow through company systems, CDOs in government could source and identify possible data sources from a range of actors. Whilst their role would be focused on partnerships and increasing the quantity of data for monitoring sustainable development, the chief statistician's role and the NSO's would be data quality; ensuring the adoption of standards, the implementation of methodologies, supervising the production of official statistics and ensuring the integrity of the results (particularly when official data is combined with new sources) to ensure relevant statistics are available to meet the objectives of the government. The CDO can also help the government to take a long-term view, by commissioning modeling work and scenario planning to prepare for the acute and unpredictable challenges of sustainable development. Appointing a high-level CDO, ideally situated within the Executive Office or a senior Ministry, also has the advantage of raising the political profile of SDG monitoring, and can ensure closer alignment to the Executive Office, whilst the NSO retains political independence and neutrality (see Table 3 on their respective roles and responsibilities).

High-level taskforces or commissions, run out of the executive office, are another successful modality for raising the political profile of data and statistics within government. These have proven particularly effective at advancing the case for open data. In Mexico, for example, the National Institute of Geography and Statistics (INEGI) partnered with the Office of the President of Mexico to set up an Open Data Technical Committee, which sought to align national statistical plans with new open data policies.[30] The committee identified more than 30 high-value datasets that could be made publicly available as part of its Strategic Open

TABLE 3: ROLES AND RESPONSIBILITIES FOR CHIEF DATA OFFICERS AND NATIONAL STATISTICAL OFFICES (NSOS)

	Chief Data Officer	Chief Statistician and NSO
Primary role	To broker new partnerships to produce, clean, compile and analyze data and produce official statistics.	To manage the impartial production of official statistics.
Responsibilities	To identify new data partners and new data sources for the government (ministries and departments) and the NSO. Broker the partnerships – including overseeing legal partnership agreements. Conduct internal advocacy to ensure the government maintains a strong spotlight on data for sustainable development, makes their data openly available, and is using an evidence-based approach to policy and decision-making.	Production of official statistics – including data on social, economic and environmental conditions, as well as national accounts. Coordination and oversight of agreed data partners. Data quality assurance, testing and evaluation.
Expertise	Multi-stakeholder partnership agreements, familiarity with both official and nonofficial data sources.	Statistical methods, to tertiary degree level.
Reports to	The executive of government. A political appointment.	N/A – produces data for government, but is administratively independent.

Data Infrastructure, including the National Statistical Directory of Economic Units, a birth registry and the Ozone Depleting Substances Project Protocol. Importantly, they opened access to socio-economic data and geospatial data, free of charge. They even went one step further, creating a national platform — agenda2030.mx — to enable the general public to use the data for SDG monitoring with a public SDG Open Data API at its core.[31]

MAINTAINING HIGH-LEVEL POLITICAL INTEREST IN THE DATA REVOLUTION AND "LEAVING NO ONE BEHIND"

By committing to the 2030 Agenda, 193 Member States made an explicit commitment to invest in data, and use it to monitor and guide their progress on the SDGs. Furthermore, governments pledged to use data to ensure we "leave no one behind" and that we track the situation of disadvantaged groups to the greatest extent possible. However, these agreements were not accompanied by firm financial targets for data investments, nor by the creation of any kind of joint financing facility, beyond the modest expansion of existing World Bank mechanisms that support statistical infrastructure development. These commitments were also not expressed as targets. The two data-related targets featured in the framework (SDG Targets 17.18[32] and 17.19[33]) address capacity-building in developing countries and identifying measures that move beyond GDP.

Despite the lack of data-oriented SDG targets or explicit financial commitments, data has become an important theme for many countries working through the process of getting started on the SDGs, as the first two rounds of Voluntary National Reviews on the SDGs serve to demonstrate.[34] However, investments in data are referred to as a technocratic solution, not as a political imperative. Data is the gateway through which the goals will be achieved; without information on resource flows, service efficiency and coverage, as well as the wellbeing of the poorest and most marginalized, the goals cannot be attained. It is therefore imperative that the international community strengthens high-level attention to the data revolution for sustainable development, as well as attention to the commitment to "leave no one behind."

One practical means by which to focus high-level political attention and to encourage the flow of resources is to establish a Secretary-General's High-level Panel or Advisory Board – with both expert and Heads of State participation ¬– focused on the data revolution and the commitment to "leave no one behind." Such a panel would be responsible for defining data investments; identifying new financial resources; and focusing energies and attention on crucial issues like the gaps in gender poverty data, the urgent need to establish a population baseline and the need to support the 2020 census round. Their activities would align with the work of the UN Statistical Commission's High-level Group (which is comprised of eminent statisticians), but would also lift it out of the technical realm and make data a political imperative, encouraging the prioritization of data investments at the highest levels of government.

Building modern data systems sufficiently comprehensive to support the implementation of the SDGs is an urgent imperative. These are the kinds of institutional structures that need to be in place from the get-go. It would therefore seem practical that such a High-level Panel or Group be established in the first few years of implementation to help put in place the building blocks for successful implementation.

MULTI-STAKEHOLDER SPACES FOR COLLABORATION

In addition to high-level political mechanisms and the creation of CDOs, good governance of the data revolution for sustainable development will require the creation of open, equal platforms for collaboration. Governments can create such a platform by inviting nongovernmental stakeholders into their SDG dialogues to review, analyze and catalogue available data together. The GPSDD – a global multi-stakeholder network looking to harness the data revolution for sustainable development – has undertaken to instigate such processes in partnership with governments via a series of SDG data forums and roadmap exercises. According to the Government of Ghana, which recently hosted a GPSDD-sponsored SDG data forum, the forum "intends to bring together a diverse set of data producers and users, as well as innovators in the field to discuss the way forward, including through exploring how new technologies and approaches can be used to strengthen the data ecosystem."[35] In this context, the data ecosystem refers to the NSO and other government offices at national and subnational levels conducting statistical activities and mapping to produce

geospatial data and Earth observations, as well as non-state actors who produce and use data. It encompasses not only official statistics, but also new technologies for enhancing participation in the statistical cycle; technological infrastructure for improved collection and use of data; and legal, policy and data quality frameworks.[36]

At the international level, the Statistical Commission recently hosted the UN World Data Forum – a space for governments, civil society organizations and private companies to share new and alternative approaches to data collection and monitoring, with particular emphasis on the monitoring of the new SDGs. These kinds of informal spaces are already helping to match demand and supply, enabling countries to articulate what they want to monitor and to invite nongovernmental actors to help them fill gaps. One such partnership, which sprung out of the GPSDD and its various multi-stakeholder meetings, is between the NSOs of Colombia and Senegal, the telecommunications provider Orange, nongovernmental organizations (NGOs) and research networks including Data Pop Alliance and SDSN. This project aims to develop open algorithms that private companies can use to share relevant sustainable development data in a way that both respects the data privacy of their customers and clients and is useful for the national statistical office.[37]

MULTI-STAKEHOLDER ACCOUNTABILITY FRAMEWORKS

An important outcome of multi-stakeholder dialogues is a public discussion of roles and responsibilities. NSOs need to invite nongovernmental partners to co-create a monitoring strategy and framework that clearly articulates what each actor will do, thereby putting in place an accountability framework. Annual contributions could then be assessed by the coordinating NSO, as well as another arm of government such as a legislative assembly that could provide a very public assessment of performance. Some form of formalized accountability framework is essential if private, nongovernmental actors are being invited to contribute data for public SDG monitoring annually and over an extended time frame. Likewise, having nongovernmental actors involved in a discussion about what the government will monitor can help to ensure governments fulfill their commitments.

At the international level, the UN World Data Forum and the GPSDD are already helping to catalyze new partnerships between government and nongovernmental actors using informal networking approaches. However, as is the case at the national level, accountability is crucial. If nongovernmental actors are going to contribute to formal SDG monitoring in a systemic way and over a long time frame, then – like government actors – they need a formal space in which to explain their activities, present their data and be vetted by the international community. The UN Statistical Commission, established in 1947, is the highest body of the global statistical system. It brings together Chief Statisticians from over 193 Member States to set statistical standards, develop concepts and methods, and design strategies for their implementation at the national and international level. It is complemented by the United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM), which is a Member State-led committee focused on setting the agenda for the development of global geospatial information. If private and nongovernmental data producers are to play a more formal role in SDG monitoring, it would therefore be reasonable to create a space for them to engage with the Statistical Commission and the UN-GGIM.

Recent statistical commissions have featured a large number of side events hosted in partnership with nongovernmental actors. These are promising signs. Furthermore, civil society groups that currently have consultative status with the United Nations Economic and Social Commission (ECOSOC) are able to participate in the annual meeting of the Statistical Commission as observers, sitting in the gallery.[38] These are welcome developments that give nongovernmental actors the opportunity to learn from and engage with the formal governmental process, but they fall short of providing a platform for equal engagement or discussion. They also do not provide any kind of accountability for nongovernmental actors contributing data to the SDG effort. **To show willingness to engage with non-governmental actors and to create a forum for meaningful accountability the UNSC should expand its annual meeting to include a dedicated session with non-governmental actors and experts.** Organizations wanting to contribute to the SDG effort would use this session to showcase their data and their methodologies, explain how it can contribute to national and global monitoring, and open their methods up to

public scrutiny. This should complement and build upon more concerted, regular nongovernmental engagement through the IAEG-SDG. At present, the IAEG-SDG is a Member State-only group, with limited consultation with UN-specialized agencies, academia and private or NGO participants. **The IAEG-SDG should open its membership to include a representative sample of vetted and approved academic, multilateral, private and NGO sector representatives with full voting rights.** It should empower these members as representatives of their community to hold annual consultations with the private and NGO sectors through which a broader array of voices is represented in the IAEG-SDG and at the annual session of the Statistical Commission.

B. PRINCIPLES AND STANDARDS

PRINCIPLES

Institutional mechanisms that facilitate data partnerships are essential, but only if all of the actors around the table speak the same language. Without internationally agreed terminology, countries can use wildly different methodologies and achieve very different results. Controlled vocabularies are an essential component of technical data standards, as they provide a precise and agreed definition of what is being measured or counted. For example, the term 'affected' within a disaster risk reduction context might have a different meaning based on an individual countries classification of who is directly or indirectly affected. This can impact on the type of response given by government and non-government agencies and as such, influence the data, how it's collected and analyzed.

Common principles and standards are therefore required to help ensure a certain quality and format of data and methods before diverse actors enter into new data partnerships. The Secretary-General's Independent Expert Advisory Group on the Data Revolution (IEAG) suggested that this common language be achieved through a new global consensus or "Compact" on principles and standards, which would "facilitate openness and information sharing, and protect human rights."[39] They highlighted nine core principles that should be common to all actors contributing data to the measurement of sustainable development. These include:

1. **Data quality and integrity:** The entire process of data design, collection, analysis and dissemination needs to be demonstrably of high quality and integrity. Clear standards need to be developed to safeguard quality, drawing on the Fundamental Principles of Official Statistics and the work of independent third parties.

2. **Data disaggregation:** To the extent possible and with due safeguards for individual privacy and data quality, data should be disaggregated across many dimensions, such as geography, wealth, disability, sex, gender and age.

3. **Data timeliness:** Standards should be tightened and technology leveraged to reduce the time between the design of data collection and publication.

4. **Data transparency and openness:** All data on public matters and/or funded by public funds, including data produced by the private sector, should be made public and "open by default," with narrow exemptions for genuine security or privacy concerns.

5. **Data usability and curation:** The data architecture should therefore place great emphasis on user-centered design and user-friendly interfaces.

6. **Protection and privacy:** Clear international norms and standards and robust national policy and legal frameworks need to be developed that regulate opt-in and opt-out, data mining, confidentiality, inadvertent disclosure, use, re-use for other purpose, transfer and dissemination.

7. **Data governance and independence:** Data quality should be protected and improved by strengthening NSOs and ensuring they are functionally autonomous, independent of sector ministries and political influence.

8. **Data resources and capacity:** Investments should be made in human capital, new technology, infrastructure, geospatial data and management systems in both governmental and independent systems, as well as information intermediaries.

9. **(Human) data rights:** Any legal or regulatory mechanisms, networks or partnerships set up

to mobilize the data revolution for sustainable development should: have the protection of human rights as a core part of their activities; specify who is responsible for upholding those rights; and support the protection, respect and fulfillment of human rights.

At the first multi-stakeholder UNSC, agree on a set of principles as part of a new global Data Compact. Invite private companies, NGOs and research institutions to become signatories to the Compact. In signing the Compact these institutions commit to respect the principles established and to be held accountable to them at each annual meeting of the UNSC.

STANDARDS

These principles provide a useful platform upon which diverse actors can share data, ideas and results. But to build full confidence and enable information exchange between governments, private companies, universities and other entities (who frequently distrust one another other), it is necessary to be more granular establishing certain standards to ensure data quality (as per Principle 1 above). Some of the standards required relate to seemingly innocuous subjects, like defining statistical boundaries (e.g. what constitutes an 'urban area,' as there are many different definitions used at present); others might relate to methodological approaches, such as the size of the sample or whether it is representative. Table 4 presents a snapshot of the range of issues relating to data design, collection, analysis and dissemination that would benefit from common standards.

Without dedicated global and national entities governing the activities of both government and non-governmental data producers, it will be hard to establish widely adopted and regulated standards. **The UNSC should, concurrent to the agreeing of principles, establish a committee to develop detailed standards to ensure data integrity across public and private actors. NSOs should mirror these processes at the national level.** A UNSC Friends of the Chair group is also taking useful first steps, by looking at possible revisions to the Fundamental Principles of Official Statistics to better cover non-official data. This is in direct response to the Cape Town Global Action Plan's recommendation that there be new guidelines on the use of innovative data generated outside the official statistical system as part of official statistics (Objective 2.3[40]).

TABLE 4: TOPICS REQUIRING COMMON STANDARDS BETWEEN PUBLIC AND PRIVATE DATA PRODUCERS TO ENSURE DATA QUALITY AND INTEGRITY

Design • Common statistical definitions and collection units • Sample sizes • Levels of disaggregation (including the five categories recognized in Agenda 2030: sex, age, geography, wealth and disability) • New technology to reduce costs and speed up production.	**Collection** • Regularity of collection • Timely publication of data soon after collection • Security and confidentiality of the microdata • Opt-in and opt-out measures • Cost efficiency • Common documentation standards such as Data Documentation Initiative (DDI) • Advance Release Calendars
Analysis • Open data standards to enable public analysis (data producer as a signatory of the Open Data Charter) • Open algorithms to enable access to private company data • Microdata held in possession of the data producer, not transferred for reuse without prior public consent • Data quality assessment frameworks: what are the recommended features and how the assessment process is done. • Analytical methods, tested and proven, are documented and shared.	**Dissemination** • Aggregate data is openly available to general public • Data is shared in an accessible, user-friendly format • Data related to specific SDG indicators is reported to a national statistical entity on an annual basis (NSO/CDO/national review board) and where relevant to an international entity (UN Statistical Commission) • Common dissemination standards, like Statistical Data and Metadata eXchange • Data use and user feedback are monitored.

In addition, the Statistical Commission should encourage the formation of new sets of "City Groups": thematic taskforces, which meet in specific cities, to tackle specific issues or consider recommend methodologies for monitoring clusters of related SDGs[41]. These groups should encourage private and nongovernmental actors to contribute to standards relevant to the particular types of data they can contribute. For example, discussions on SDG 3 (relating to good health and well-being) might invite the participation of private healthcare companies that are tracking morbidity and disease burden, or academic institutions that are conducting surveys or analyzing administrative data to better understand disease patterns and distribution. These epistemic communities should work together to develop sector-specific standards relating to the production, analysis and sharing of population and health data – for example, establishing standards around the privacy and anonymity of microdata. Discussions on SDG 14 (relating to terrestrial ecosystems) will necessitate a very different set of actors and standards. Actors might include private companies such as satellite operators or commercial farmers or fisheries, as well as scientific research units and environmental civil society groups, while required standards might relate to ways of measuring biodiversity or feasible levels of geographic disaggregation. There have been a number of successful city groups established in the past. Two examples are the Delhi Group (established 1997), which helped improve measurement of the informal sector, and the Washington Group (established 2001), which has helped to standardize the measurement of disability within censuses and surveys.

Taking a sectoral approach to establishing data standards at the national level and/or under the auspices of UN Statistical Commission should help to unite epistemic communities in their quest to monitor specific SDGs and make the process of establishing standards easier to convene and manage. It is worth noting, however, that standards in crosscutting issues like sampling and quality assurance will require a multi-sectoral and multi-disciplinary approach.

BOX 5: STANDARD SETTING FOR IMPROVED INTEROPERABILITY OF EARTH OBSERVATION DATA

The intergovernmental Group on Earth Observations (GEO) is leading an effort to build a Global Earth Observation System of Systems (GEOSS). GEOSS builds upon, and adds value to, functioning Earth observation systems by supporting their interoperability, among other objectives. The added value primarily comes from interoperability facilitating the creation of datasets from disparate observation systems for space, land, sea and atmospheric data that can be used to obtain vital information for the benefit of society. Interoperability in GEOSS is achieved primarily by specifying how GEOSS components exchange data and information at their interfaces, including use of an up-to-date Discovery and Access Broker (GEO DAB) utilizing common API formats. The GEOSS strategy is to realize a system of systems through adoption of a broad range of international standards that enable interoperability. The GEOSS Common Infrastructure (GCI) offers approximately 400 million open Earth observations data and information resources via the GEO DAB or online at the GEOSS Portal.

PROTOCOLS FOR DATA STORAGE AND SHARING

Some of the most difficult issues on which to establish standards or protocols relate to how, why, when and by whom data are stored and shared. As the boundaries of what is possible technologically continue to expand, the ethical landscape will be increasingly challenging to navigate – and the growing sophistication of some stakeholders will further intensify the digital divide. Biometric data, for example (which is increasingly used by governments to provide unique identifiers, improve birth registration, and access to services), has huge potential but is also highly sensitive. In the wrong hands, or in the hands of an entity without sufficient capacity to safely store and manage the data, it could be exploited easily for commercial gain or to suit political ends.

The storage of data records is a crucial issue, often overlooked. NSOs are commonly subject to clear rules and mandates regarding the storage and protection of original census and survey records – including highly sensitive individualized micro data – and other government departments hold equally confidential administrative records. If nongovernmental actors are to contribute data to monitor sustainable development, they too need to have high ethical standards relating to data storage, protection and upkeep. In more than 100 countries worldwide, data protection acts help to ensure that data held by private companies are subject to the same protections as that held by governments.[42] However, that leaves more than 90 countries worldwide without effective mechanisms to ensure data protection. Inevitably it is these countries with the weakest regulatory frameworks that are most in need of human, technological and financial support from third parties both to generate and to retain and safely store data. **International agencies should support low- and lower-middle-income countries to put in place essential data protection safeguards like data protection laws and acts.**

With assistance from international agencies, governments should also establish standards and mechanisms for private partnerships. **International agencies like the UN, World Bank and OECD can play a role here too, helping low-capacity countries to establish strong legal and regulatory data frameworks within which non-governmental actors should operate.** These frameworks should lay the foundations for equal partnerships between governments and nongovernmental actors and ensure that, no matter who is producing the data, the public can be assured that their personal data are being protected and effectively stored over time. Privacy Commissions can are a helpful mechanism when trying to utilize private data for the public good, ensuring adherence to legal arrangements and frameworks, but with sufficient flexibility so as to enable the use of private data for clear public purposes like the monitoring of the SDGs.

International agencies should also work with new global initiatives (for example, new initiatives that aim to make available higher resolution satellite imagery) to ensure that these well-meaning endeavors also maintain high standards for data protection and storage – sharing data openly but with a clear understanding and measure of its intended use.

The complex web of concerns relating to the privacy and security of highly-personalized microdata is brought into sharp relief when data is shared between partners. But using private data to help build a modern data system and to fill data gaps depends upon data sharing – so how can it be done safely?

To identify the best methods and appropriate standards for data sharing, one first needs to consider why the data is being shared and the associated risks with each scenario. For example, is data being shared to monitor an issue more frequently, with greater speed or for less money? Or is it being shared in the name of science and discovery – e.g. to create a large data platform that enables us to analyze, interpret and discover meaning, rather than gathering data to test or monitor preconceived ideas?

For the former, responses might entail bilateral sharing agreements, multilateral sharing platforms with common standards like Humanitarian Data Exchange (HDX) or more technological solutions like end-to-end encryption services (see section C below). The latter scenario requires a completely different approach, in which one must think through how to create a 'clean room' in which data scientists can freely explore data without constraints, but without risk of that data being shared more widely. Both of these methods require standards and protocols, but they are highly context-specific. International organizations like the GPSDD can share replicable best practices with regards to bilateral and multilateral agreements and/or platforms. They should also solicit the advice of lawyers and private industries working with highly personal data, such as insurance and risk management, to learn best practices and appropriate codes of conduct. At the national level, CDOs – with the mandate for identifying and safely integrating new data sources into the official statistical system – should work with the NSO, other government departments and third parties to design data-sharing agreements or conditions that are tailored to specific local needs. In this way, they ensure the greatest possible protection of individual data. CDOs should also help the relevant parties think through the potential adverse consequences of their data sharing – taking a user's and subject's perspective to think through the benefits, risks and trade-offs.

BOX 6: SHARING DATA NATIONAL, SUBNATIONAL, PUBLIC AND PRIVATE ACTORS IN BOGOTÁ, COLOMBIA

In 2016, the Sustainable Development Solutions Network (SDSN) in alliance with the Centro de Pensamiento Estratégico Internacional (CEPEI) launched an initiative to reconcile data from subnational levels of government with national data being used to monitor the SDGs. Launching the project was dependent upon identifying a subnational data producer that was willing to share their data with the National Administrative Department of Statistics (DANE), the Colombian NSO in charge of official statistics and progress reporting on the SDGs. In late 2015, the Bogota Chamber of Commerce was identified as an appropriate partner; it had data that could help DANE monitor SDGs 8 and 9 (on economic growth, and infrastructure and industrialization, respectively).

The Colombian Chamber of Commerce (CoCB) is an organization that brings together formal companies under Colombian law and obliges them to register with Chambers of Commerce in the different cities in which they operate. In accordance with Colombian law, the CoCB asks companies for the mercantile registry form in which they must register the economic activity of the company from the previous year. The CoCB also conducts surveys and keeps administrative records that enable the tracking of the businesses, new enterprise and economic growth.

The collaboration between DANE and the CoCB has revolved around a process of data reconciliation, which aims to ensure data from the CoCB complies with the statistical quality required by NSOs. The data reconciliation process has involved four steps:

1. awareness-raising of the aims and ambitions of the project with DANE staff and the Bogotá Chamber of Commerce;

2. definition of databases;

3. definition of indicators; and

4. quality analysis of the mercantile registry to find out if it meets the quality standards required by DANE, so that the information can be used to measure six indicators relating to SDGs 8 and 9.

An ongoing, crucial step has also been establishing a legal data-sharing agreement enabling DANE staff to have access to the CoCB database. This has been a time-intensive process, but crucial to the longevity of the project and future collaborations between DANE and the CoCB.

One year in, the data quality analysis has revealed that the mercantile registry data does not fulfill the quality standards set by DANE. DANE has therefore recommended that the CoCB improve its rules of validation and consistency before transmitting data to DANE in order to obtain more accurate information on the mercantile registry questionnaire. The project has also:

- established a methodology to access private sector data and carry out quality analysis;

- encouraged the CoCB to conduct programming on “R,” (an open source programming language for statistical computing) which will help the CoCB to improve quality of data to measure SDGs at the subnational level; and

- documented various lessons learned for data sharing among partners. These lessons include the importance of:

    - trust-building among partners before embarking upon technical exchange

    - support from senior management for quick and practical implementation of the established actions

    - generic institutional protocols that flag things, such as when legal review will be required

    - explaining the meaning and relevance of SDGs to kick-start a among discussion stakeholders

Source: Written by Fredy Rodriguez, Centro de Pensamiento Estratégico Internacional (CEPEI)

C. TECHNOLOGY, INNOVATION AND CITIZEN-LED ANALYSIS

NEW APPROACHES TO DATA SHARING

Thanks to this era of constant technological innovation, there is a range of newly created tools and safeguards that companies and governments can employ to protect the privacy of the data being shared between two parties and lower the transaction costs and human capacity requirements associated with standardizing and sharing data safely. For example, some companies have come up with end-to-end encryption services that only the sender and recipient can access. Companies like WhatsApp and Signal are using these approaches. These methods allow the secure exchange of data, but do not attempt to ensure the protection of the data source. Another approach is to use 'data enclaves' where researchers or government statisticians gain access behind a firewall to run their analyses without being able to extract confidential information. When the West African Ebola virus epidemic (2013-2016) struck, local mobile telecom providers enabled the WHO and various other front-line service providers to look at population movement and call patterns to assess disease spread.[43] These analyses were done within 'data enclaves,' where specific data sets were cleaned and anonymized by the private company before being shared with analysts who had temporary access to the files. This kind of approach works for piecemeal analyses, but does not enable regular or large-scale data exchange. To overcome this problem a number of entities are trialing the use of microdata labs; virtual or physical facilities which enable broad access to secure, sensitive, detailed data for approved researchers working on defined and approved non-commercial projects, which serve the public good. The UK's Office for National Statistics has a virtual microdata laboratory, as does Germany (the German Microdata Lab), which offers services for microdata from Eurostat (European Microdata) and engages in methodological and substantive research based on microdata from official statistics in general. These kinds of laboratories lower the cost for individual researchers in obtaining information and enables one central entity to ensure safe, well documented sharing of information.[44] At present microdata labs are predominantly used by NSOs, but moving forward they should be broadened to include all government data, and should be made accessible to academics, CSOs and other approved, third party data providers and users to encourage widespread analysis and data modeling.

Orange, the mobile and Internet service provider, is also trialing a new approach to large-scale proprietorial data sharing. Orange is teaming up with a range of research organizations (including SDSN) and government partners to develop a series of certified, open algorithms through the Open Algorithms (OPAL) Project. With these algorithms, companies can run regular analyses behind their firewall and share the data in larger quantifies and more regularly with the public. For regular, systematic sharing of private company data with NSOs and other public entities, mechanisms such as these are going to be essential. **Private companies and the international community should work together to encourage data-sharing innovations by organizing challenges or competitions, for example a data sharing challenge at Davos or the 2019 World Data Forum.**

INTEROPERABILITY

Data interoperability is one of the biggest barriers to effective public use of private data – particularly with regards to disaggregation, as data need to be in a comparable format and/or use comparable standards if they are to be overlaid or combined. The open data movement has made good headway in encouraging public and private actors to work together to publish and share data using comparable formats.[45]

For example, the Blue Button initiative is a "partnership between the [United States] health care industry and the Federal Government that aims to empower all Americans with access to their own electronic health information." The initiative began by making individual medical data from the U.S. Department of Veterans Affairs available to vets, but has been expanded to serve a much wider constituency. The initiative makes it easier for citizens to access their health records and facilitates sharing of those records among various medical providers.[46]

Another excellent example of a public-private partnership for improved data availability and granularity using a set of common interoperable standards is provided by the Partnership for Resilience & Preparedness (PREP) platform for tackling climate change.[47] PREP is a public-private initiative with the U.S. Department of State, local governments and a number of private partners. This platform allows

city entities to create and embed a climate data portal for their individual municipality and populate it with global datasets from multiple space agencies trimmed to their geographic boundaries. Cities can then upload their own locally-generated data to create a more robust, accurate and ultimately actionable picture for local resilience than any one of those datasets alone could offer. It is a particularly important example of local actors being able to access free, high quality geospatial imagery which is often costly to purchase. This reinforces the need for collaboration and cooperation.

Developing common standards is highly sectoral and specific to each public and private partnership. But according to the Joined-up Data Standards initiative (JUDS), there are five common principles to help guide public and private actors seeking to make their data more readily available and useable by third parties, here elaborated by SDSN:[48]

1. **Use existing data standards:** Where possible, those seeking to develop a new standard should spend time considering what is already out there and whether an open data standard already exists that can simply and easily be adapted to their needs. If a new standard is absolutely necessary, then it should be compatible with existing standards.

2. **Don't forget metadata:** Metadata includes information on the source of a piece of data, its author, the version being published and the link to the original dataset. It provides all relevant information required to contextualize data, and therefore also explains the reliability of the data. Metadata also needs to be made available in standardized formats that are machine-readable.

3. **Use common classifications:** Language and classifications about data need to be the same. Even the slightest difference in definition can hinder machine-readability and interoperability of data; "The identify-org.net site succinctly explains why the issue of organizational identifiers is important: "If my dataset tells you I have contacts with 'IBM Ltd'. 'International Business Machines' and 'I.B.M' – how many firms am I working with?" Unique identifiers would go a long way to overcoming basic semantic challenges like this." For the international development community, a comprehensive inventory of all international and relevant national classification systems and standards could go a long way to improving this problem.

4. **Publish data in machine-readable formats:** For ease of use, efficiency and scale data should be made available in machine-readable formats such as Resource Description Framework (RDF), eXtensible Markup Language (XML) and JavaScript Object Notation (JSON). Use of these formats would enable a computer to access, identify and filter data in an automated way. However, machine readability should be secondary to human readability. Data also needs to be available in formats that humans can easily manipulate, such as comma-separated vales (CSV) or Microsoft Excel Open XML Format Spreadsheet (XLSX).

5. **Ensure standards are user-driven:** Open (and theoretically interoperable) data does not always equate to useable data, as for data to be useable it needs to be presented and designed with the user needs in mind. JUDS provides the example of the Humanitarian eXchange Language (HXL) standard that is accessible and functional because of its incredible simplicity and ease of use. The standard was developed through an iterative process with users, figuring out what was most accessible for the user and taking into account what they needed to do with the data. The same could be said of collaborative project OpenStreetMap (which is increasingly used to map informal areas and humanitarian contexts), or of the standards developed by the Open Geospatial Consortium, which use a consensus-driven process to develop common standards designed by a wide range of stakeholders within any given sector. See, for example, their ongoing effort to create an Arctic Spatial Data Infrastructure among the eight Arctic countries' national mapping agencies.

To increase the amount of data available for disaggregated monitoring, the international community and national governments need to promote these principles for interoperability. **UNSD should update and identify gaps in their UN Classifications Registry to include classification systems being used by other international and large-scale epistemic communities, as well as relevant national systems, thereby making available common standards and registries for all entities looking to make their data interoperable.**

CITIZEN-GENERATED DATA

Citizen-generated data (CGD) is data that people or their organizations produce to directly monitor, demand or drive change on issues that affect them. According to one leading initiative in the sector, DataShift, "[CGD] can be produced through crowdsourcing mechanisms or citizen reporting initiatives, often organized and managed by civil society groups."[49] This approach received frequent discussion in the negotiations on the post-2015 development agenda, thanks to the work being done by organizations like DataShift and the Humanitarian OpenStreetMap Team (HOT) to legitimize the work of citizen and civil society data. DataShift is an initiative that builds the capacity and confidence of civil society organizations to produce and use citizen-generated data to monitor sustainable development progress, demand accountability and campaign for transformative change. At the Pulse Lab Jakarta in Indonesia, citizen-generated data is being mined to better inform policy decisions. "Mining Citizen Feedback Data for Enhanced Local Government Decision-Making" combines various sources of citizen-generated data — from local and national complaint systems to informal Twitter comments — to better understand local conditions and citizen concerns:

"The results demonstrate the potential utility of (a) near real-time information on public policy issues and their corresponding locations within defined constituencies, (b) enhanced data analysis for prioritization and rapid response, and (c) deriving insights on different aspects of citizen feedback. The publication of citizen feedback on public-facing dashboards can enhance transparency and help constituents understand how their feedback is processed."[50]

Another good example of the utility of citizen-generated data comes from Dar Ramani Huria, a community-based mapping project in Dar Es Salaam, Tanzania. Dar Ramani Huria (Swahili for "Dar Open Map") is training teams of local university students and community members from throughout Dar Es Salaam to use OpenStreetMap to create sophisticated and highly-accurate maps of Dar es Salaam. These neighborhoods, known as wards, were selected because they are the most flood-prone areas of the city. By helping communities to map residential areas, roads, streams, floodplains and other relevant features, the project aims to bring disaster prevention and response to areas that were previously off the map. The project also brings awareness of the need for flood prevention and risk reduction to the local level, while teaching participants valuable computer and mapping skills that they can put to use elsewhere. Since 2015, Dar Ramani Huria has mapped 29 wards, including 1,254 kilometers of waterways and 3,396 roads. It has also trained over 450 community mappers.[51]

Current use of CGD is limited, however, by the lack of common standards and methods, which discourages those in the official statistical community from using it. For all the excellent, large-scale CGD efforts being run by groups like DataShift and the Humanitarian OpenStreetMap Team, there are a large number of small-scale, relatively unsophisticated citizen consultation exercises that claim to be CGD, but don't have the sample size or technical rigor to be useful for official monitoring. To make CGD an integral part of the SDG monitoring process, the CGD community should work to establish a methods board or community of practice that can help to establish standards and parameters for robust CGD — working not only with CGD practitioners, but also official statisticians from NSOs. **The newly created GPSDD working group on CGD should look to establish an inter-agency and expert group on CGD (or a "city-group") which can help to set standards and common methods for CGD to encourage greater uptake of CGD by NSOs. The group should promote the creation of CDOs within government who can help ensure a steady stream of high-quality CGD is being fed into the national data collection process.**

D. CAPACITY AND RESOURCES

STATISTICAL CAPACITY

The SDGs ask governments to monitor a broader and more complex set of challenges than ever before, and to do so with greater granularity and specificity. For the most part, the task has been laid at the feet of NSOs, which have historically been sidelined and underfunded. But as the complexity of governance increases, with increased environmental insecurity, growing inequalities and larger, more transient populations, data will be crucial for evidence-based planning. To which end, NSOs need to evolve into broader, cross-sectoral and cross-departmental systems with greater capacity and resourcing (as described under Pathway A).

Statistical capacity is defined as "the sustainable ability of countries to meet user (government, policy makers, researchers, citizens, business) needs for high-quality data and statistics (i.e. timely, reliable, accessible, relevant)."[52] It can be defined on four levels: human, technical, financial and organizational. (Financial and organizational capacity is discussed in more detail in the section entitled "Financing data investments" below.) Building the human and technical capacity of government departments and the NSO will be achieved through three mechanisms: education and training, technical partnerships and technological exchange.

- **Education and training:** College-level and mid-career statistical training programs should be prioritized to help develop and maintain a cadre of qualified statisticians within government, well versed in classical statistical methods, GIS and partnership models for data collection. The international community can support countries with limited capacity by creating open and free educational materials like massive open online courses (MOOCs), by hosting technical trainings for NSO staff and national statisticians and by providing opportunities for remote learning and e-certification. A good example is a technical seminar series hosted by the Data Pop Alliance and UN Systems Staff College in partnership with the Hewlett Foundation, aiming to enhance participants' 'big data literacy.' In addition to technical skills the new approaches to capacity development have also started to promote building soft skills including management, leadership, and risk management.

- **Technical partnerships:** Developing technical competencies in the wide range of data science approaches can be achieved through formalized training, but also through technical exchange. Peer-to-peer learning can be a vital tool when trying to foster partnerships between diverse actors, particularly across the public and private sectors. Examples of technical peer-to-peer learning programs include the American Statistical Association's Statisticians Without Borders program and recent efforts by the GPSDD to bring private actors with specific expertise into national SDG roadmap exercises. PARIS21 – the Partnership in Statistics for Development in the 21st Century - also attempts to foster cross-sectoral and multi-stakeholder data partnerships through the modernization of countries' National Strategies for the Development of Statistics (NSDS).[53] These efforts are important, but due to funding constraints they are piecemeal. The Statistical Commission and GPSDD should work with public and private actors organized by epistemic communities to develop technical training materials for NSOs and for private data partners. With this information, they can better understand each other's respective skills and capacities before embarking on joint monitoring or data-sharing exercises. By way of example, telecommunications companies should develop technical materials on their data collection approaches, storage capacities, metadata and so on that can be studied by NSOs looking to utilize this data — to assess human mobility, for example. In exchange, NSOs should share the methods they employ to measure population density, movement and livelihoods so that both groups can learn from each other's methods and best understand how to overlay their data. Likewise, as a matter of principle, academic partners should develop open, publicly-accessible training resources relating to their analytical methods before utilizing national government data sets to conduct their third-party assessments.

- **Technological exchange:** For capacity-constrained countries, information and communications technologies (ICTs) have huge potential to speed up data collection and to improve technical capacities. By way of example: In 2002, Senegal took four years to finalize their census. In 2013, the country turned out its results in one year. They achieved this by using personal data assistants (PDAs), brought in through cooperation with the Brazilian Institute of Geography and Statistics (IBGE). The IBGE provided hardware in the form of PDAs and training in the use of them.[54] Technological and training partnerships such as this can help low-capacity countries leapfrog, building up a strong statistical baseline in a comparatively fast rate and with relatively low capacity. The World Bank and GPSDD data innovation trust fund[55] should prioritize these kinds of technological exchanges, which have the potential to fill crucial data capacity gaps, improve the coverage of data and aid in its timeliness.

Ultimately, it is national governments that will decide their capacity requirements. These requirements are often laid out in NSDS or, more recently, in SDG roadmaps. These strategies should serve as a touchstone for third parties looking to provide support so that their assistance matches demand and speaks to local, contextual priorities. Groups like PARIS21 do a huge amount to support countries to undertake NSDSs. PARIS21 is also coordinating a timely initiative to look into new approaches needed to boost capacity development and what needs to change. But, there is an acute need for more widespread support. **Working with PARIS21 and the GPSDD, the UN Statistical Commission and UN system as a whole should promote SDG data roadmaps, which articulate the functions of the NSO, CDO and other local data partners including academia, private companies and NGOs, in all LIC countries without current, effective NSDSs or work to align the NSDS with SDG-related data requirements.** They should facilitate this process by promoting countries' NSDSs and by establishing innovative financing mechanisms (as per Global Action Plan Objective 6.1[56]) that respond to countries' capacity needs. For more recommendations on financing, see point iii below.

BUILDING DATA LITERACY

Whereas statistical literacy refers to one ability to understand statistics and statistical methods, data literacy relates to individuals' ability to find, understand and interpret data to make informed decisions. Data Pop Alliance has also suggested that data literacy reflects a desire to constructively engage in society through data. Whatever the definition, in an increasingly data-oriented world, competence with data and an understanding of how it intersects with our lives are crucial. Not only can data literacy facilitate more informed decisions, but it can help individuals to understand their opportunities, challenges and rights, thereby enabling them to more constructively engage with government and private companies.

Data literacy is an area of increasing attention as policy makers acknowledge that a data-literate population can be more economically productive. Concurrently, nonprofit organizations argue that data literacy can empower marginalized groups to seize more economic opportunities and to feel more entitled to demand their rights. According to Data Pop Alliance, "Schools and nonprofit organizations (such as Code for America, Girls Who Code, School of Data, etc.) have emerged to tackle the digital divide by providing data literacy and coding programs for vulnerable populations, specifically for women and minorities. An increasing number of data journalists are using and writing about data. Open data and civic technology advocates have organized hackathons for civic hackers to use technical skills and foster new conversations on data for social good."[57]

Global internet expansion (from 20.6 percent of the world's population online in 2007 to an estimated 46.1 percent in 2016[58]) also provides a unique opportunity for those who are data-literate. Where basic data literacy skills exist, there is great potential for citizens to access data, interpret it and use it to increase accountability, enabling citizens to push for change.

The Global Action Plan calls for increased data literacy via large-scale education programs. Schools are indeed the best place to foster widespread awareness of basic data concepts, collection methods, analytics and communication. **To promote data literacy on a large scale, the UNSD should establish a partnership with UNESCO and the Global Partnership for Education, through which curricula broad-based curriculum in statistical literacy could be developed for schools and learning centers around the world.** In addition, statistics training should be made available as Massive Open Online Courses (such as those developed by the SDG Academy) so that secondary- and tertiary-stage students (and those outside of the school system) can also benefit from the content. These basic data literacy resources could even be useful for mid-career professionals and executives as, according to McKinsey, even in the U.S.A. there is an acute shortage of professionals who can make data-informed decisions.[59] These introductory materials would complement more technical or specialized materials, developed for individuals with an existing grasp of statistical methods (as highlighted under "Education and training" above).

FINANCING DATA INVESTMENTS

One of the greatest barriers to monitoring sustainable development and achieving a modern data ecosystem is sustainable financing. In principle, responsibility for funding national statistics systems lies with national governments. But many countries facing urgent demands for scarce resources will not be able to finance the development of their statistical systems solely from their own budgets. The funding challenges are, of course, greatest in the poorest countries. But even wealthy country national statistics offices are not immune to the challenges of unstable and tightening budgets. The net effect of this is a tightening fiscal envelope for data and statistical capacity development just when we need it most.

A recent study of NSOs in Africa put into stark relief the significant additional challenges faced by NSOs in developing countries. The study by the Center for Global Development and the African Population and Health Research Center found that budget limitations were one of the most frequently cited reasons for lack of progress on statistical capacity in sub-Saharan countries. Of the 54 African Union member countries, only 12 are considered to have an autonomous NSO.[60] In the remaining 42 countries, statistics fall under the jurisdiction of another government ministry.[61] And in many countries, nearly all core data collection activities are funded primarily by external sources[62] — the net effect of which is investments directed by the donor, rather than based on local needs. By way of example: A recent study by the Partnership in Statistics for Development in the 21st Century (PARIS21) found that there were 3.4 data portals in Africa per country versus only 1.2 per country in Europe, reflecting donors' interests in making national data available on public platforms, rather than country input on where investment could be most usefully directed.[63]

As with other public goods, the most sustainable way to finance national statistics systems is to use domestically-generated resources from progressive taxation systems. However, in developing countries especially, increasing the tax take in a way that ensures all members of society are paying according to their means is a great challenge. Supporting developing countries to build progressive tax systems and tackling the international barriers to domestic resource mobilization by cracking down on illicit financial flows and ending the use of tax havens will be essential if we are to ensure developing countries have sufficient public resources to fund essential services and effective institutional systems, including comprehensive data systems.

But how to bridge the gap in the meantime? Estimates from SDSN and Open Data Watch suggest that developing countries (the 77 International Development Assistance, or IDA-eligible countries) will need help from donors for at least half of the estimated costs of monitoring the SDGs up to 2030,[64] and more if they are to put in place robust, integrated and interoperable administrative data systems. Investment and annual operating costs for the IDA-eligible countries were estimated to be in the order of USD 925 million. Data collection to meet additional requirements for the SDGs brings the total to USD 1.2 billion.[65] Total aid needed to support the production of the readily monitorable ("Tier 1 and Tier 2") SDG indicators is expected to be USD 635 million to 685 million a year over the period 2016 to 2030.[66] This means that to support the production of SDG indicators, an annual increase in aid of USD 350 million to 400 million is

needed. Additional resources will also be needed at the international level to support the development of methods, standards, and guidance for Tier III indicators.

Unfortunately, the trend line for official aid dedicated to supporting statistics is going in the opposite direction. A recent report by the GPSDD and Open Data Watch found that between 2015 and 2016, aid for statistics decreased by about 11 percent (see Box 7).[67] It will be necessary to continue to make the case for better data at the international level and in individual countries. The sums involved are relatively small when compared to the scale of investments needed in other development priorities, and the impact of having better data in terms of more efficient policy-making and effective identification of priorities in the individual countries significantly outweigh the costs.

One approach to aligning country and donor interests around investing in development data could be to form 'data funding compacts' among individual countries, the international community and private companies. The starting point of the compact must be the country's own data priorities, as set out in an SDG roadmap or a recent NSDS. Countries would commit to making improvements and investments in their statistical systems and data processes, and the international community would commit to providing realistic financial and technical support.

Given the pressures on international development assistance, it will be necessary to also explore other sources of funding beyond domestic taxation and aid from donors. Experience shows that the potential for raising revenue directly from statistical activities, such as undertaking work on commission and selling products, is limited. Some areas of additional financing that could be investigated, however, include corporate sponsorship and a multi-donor global fund for data[71].

Corporate sponsorship: Data-oriented companies, such as SAP Technologies, Google, and InsightSquared, have an implicit interest in supporting the development of statistical capacity of countries in which they operate. They can do this by providing financial resources, technologies, or software and expertise. With careful provisions to safeguard data privacy and independence, and under the oversight of a national chief data officer, these partnerships can teach private companies a lot about the relevance and utility of their tools for the public sector (which in and of itself provides a business opportunity), whilst also demonstrating corporate social responsibility. **The GPSDD as a consortium of public and private actors with a common interest in strengthening national statistical systems should facilitate a public-private sponsorship platform for national statistical capacity.**

A multi-donor global data fund: A data-focused multi-donor global fund, or country-level basket funds, dedicated to providing financing, capacity and technical assistance to SDG data efforts – including broad statistical capacity development - could provide the kick-start investment needed in many of the countries whose national statistics systems are furthest behind. The idea of a multi-donor fund was proposed in 2015 at the Third International Conference on Financing for Development in Addis Ababa Ethiopia[72], and resulted, in part, in some new innovation-focused financing opportunities. But a dedicated, large-scale multi donor fund could go a long way to unify global data investment priorities and to help leverage public and private resources. Care would need to be taken to ensure that a global fund for data is aligned with country road maps and NSDSs. But a lot can be learned from other successful multi-donor trust funds such as the Global Fund for AIDS, Tuberculosis and Malaria, which is structured around the development and resourcing of national health strategies.[73] **The GPSDD would be well placed to convene a dialogue among multilateral, bilateral and philanthropic donors to establish a multi-donor trust fund for statistics that consolidates and focuses the inflow of data-related resources to capacity constrained countries and statistical systems and leverages private investment for SDG monitoring.** They should identify a suitable host institution, with a proven track record in managing large financing facilities, such as the World Bank.

BOX 7: INCREASING AND RETHINKING DONOR SUPPORT TO STATISTICS

Many national statistical systems – especially those of low-income countries – depend on donor support to build statistical capacity to meet the data demands of the Sustainable Development Goals (SDGs). To meet the SDG data demands, national statistical offices – even those in high-income countries – need to upgrade and modernize their data systems. Despite the importance of accurate, timely and reliable data for improved development outcomes; the well-recognized need for modernization of official statistics; and the opportunities presented by the data revolution, current levels of official development assistance (ODA) remain inadequate.

According to the Partnership in Statistics for Development in the 21st Century (PARIS21)'s "PRESS 2016" report, the share of ODA dedicated to improving data for development was a mere 0.25 percent in 2014.[68] And it is not only the amount of financial support that is inadequate. A large share of support comes from a very small number of donors; the top five providers of ODA for statistics are responsible for 72 percent of total commitments.[69] Building a high-functioning statistical system for the SDGs and for other purposes will require additional support from the international donor community. The "State of Development Data Funding (SDDF) 2016" report by the Global Partnership for Sustainable Development Data (GPSDD) notes that an annual increase in ODA of USD 350 million to USD 400 million per year is needed to meet the data demands of the SDGs in the 77 International Development Association (IDA)/blend countries and 67 other middle-income countries[70].

Calls for increases in investment for statistics should also include discussions on how traditional donor support can be more effective. Principles such as improved donor coordination, alignment of support with national statistical plans and results-based monitoring of outcomes, among others, should guide investments for statistics. New public-private partnerships and funding mechanisms should also be explored. Increased investments that follow such a revitalized approach can change a vicious cycle of inadequate resources in the statistics sector to a virtuous one, where improved quality leads to higher use and value. If the data revolution is to deliver on its promises of sustainable development, as well as greater equality and prosperity, increased and smarter investments by donors are needed.

Source: Written by Dierdre Appel, Eric Swanson and Shaida Badiee, Open Data Watch

5. Counting on the World: A Roadmap for Urgent Action

A 21st-century data system that is fit for purpose to both monitor and achieve the SDGs will need to help us plan and prepare for the future, manage and govern more effectively, and track our progress to ensure we stay on course to meet our objectives. This requires a new approach to data and information systems that places data at the heart of government. An evidence-based approach to decision-making is vital if we are to meet the wide-ranging and ambitious SDGs within the remaining years available.

But it is not only governments who have a role to play. Private companies, universities, civil society and other third-party actors will need to contribute given the scale of the challenge. These partners can offer new skills, technologies and sources of data to improve our knowledge and understanding of sustainable development. Partnerships are not a silver bullet though. Inviting more actors into the statistical production process and using new sources of data will create organizational and methodological challenges, as well as raising important questions relating to data privacy, ownership and use.

These challenges are well known – but so, too, are the limitations of current statistical systems.[74] We need to experiment and try new partnership models if we are to produce the kind of high-quality, high-frequency data we need during the SDG period. This report aims to provide solutions. It has attempted to explain the kinds of data we need to achieve the SDGs, has identified the roles and responsibilities of different actors and profiled the urgent changes we need to make if we are to build an architecture capable of responding to the increasing demand for high-quality, disaggregated and high-resolution data.

Table 5 summarizes some of the practical steps that governments, nongovernmental organizations – including civil society groups, private companies and academia – and the international community can take to effect change now. Beside each recommendation, we identify a lead agency and provide a timeline for its fulfillment. As the recommendations provided build upon the Cape Town Global Action Plan, we call upon the Statistical Commission's High-level Group (HLG) to propose these recommendations at the UN Statistical Commission in March 2018 and take action between now and then to expand the Statistical Commission so it has full and active multi-stakeholder engagement. The HLG, GPSDD, SDSN and supportive member states should also convey relevant recommendations to the Secretary-General's office for commitment by the 2018 UN General Assembly (UNGA). SDSN TReNDS commits to monitor progress on these recommendations over the long-term, with a progress report every two years, and to continue to provide independent analysis on the state of the global data ecosystem.

Only with decisive action now will we achieve the data revolution for sustainable development, and put in place the data building blocks essential for achieving the SDGs.

TABLE 5: A ROADMAP FOR URGENT ACTION

Pathways for Action (and links to the Theory of Change)		Recommendation	Lead and/or group to facilitate this	Timeframe
Governance & Leadership – Related to theory of change priorities 1 (maintaining high-level commitment) and 3 (creating an enabling collective efforts amongst data communities).	1	Establish Chief Data Officers in all countries	Office of the DSG	By the UN General Assembly 2018
	2a	Expand the annual meeting of the UN Statistical Commission so that the official proceedings include a session for non-official data producers to showcase their data and open it up for methodological review. These sessions could be structured around individual SDGs or types of relevant data e.g. CRVS, population estimates, geospatial or earth observation measures.	Statistical Commission, with UN Statistics Division (UNSD) and High-level Group	By March 2018
	2b	Expand the membership of the IAEG-SDG to include representatives of non-governmental data producers.		
	3	Establish a Heads of State-level Taskforce or High-level Panel (akin to the High-level Panel on the Post-2015 Development Agenda) on the Data Revolution, including the theme of "leave no one behind" within the monitoring agenda.	Office of the DSG	By March 2018
	4	Establish SDG data roadmaps that articulate the functions of the National Statistical Office (NSO), Chief Data Officer (CDO) and other local data partners – including academia, private companies and nongovernmental organizations (NGOs) – in all low-income countries without current, effective national strategies for the development of statistics (NSDS) and/or work to align the NSDS with SDG-related data requirements.	National governments with GPSDD support	By December 2018
Principles & Standards – Related to theory of change priorities 2 (closing persistent data gaps), 3 (enabling collective efforts), and 4 (harnessing the data revolution)	5	At the first multi-stakeholder UNSC, agree on a set of principles as part of a new global Data Compact. Invite private companies, NGOs and research institutions to become signatories to the Compact. In signing the Compact these institutions commit to respect the principles established and to be held accountable to them at each annual meeting of the UNSC.	UNSD with the GPSDD	By March 2018
	6	Concurrent to the agreeing of principles, establish a committee to develop detailed standards to ensure data integrity across public and private actors. Mirror these processes at the national level.	UNSC	By March 2018
	7	International agencies should support lower income countries (LICs)/ low- and middle-income countries (LMICs) to put in place essential data protection safeguards like data protection laws and acts.	Statistical Commission, with the World Bank, the International Monetary Fund (IMF) and the Organization for Economic Cooperation and Development (OECD)	Ongoing
	8	International agencies like the UN, World Bank and OECD should help low-capacity countries to establish strong legal and regulatory data frameworks within which non-governmental actors should operate.	National governments	Ongoing

TABLE 5: A ROADMAP FOR URGENT ACTION (CONT)

Pathways for Action (and links to the Theory of Change)		Recommendation	Lead and/or group to facilitate this	Timeframe
Technology, Innovation & Analysis – Related to theory of change priorities 2 (closing persistent data gaps), 3 (enabling collective efforts), and 4 (harnessing the data revolution)	9	Instigate an annual challenge at Davos or the 2019 World Data Forum on data sharing innovations from private companies.	World Economic Forum (WEF) and GPSDD	By January 2018
	10	UNSD should update and identify gaps in their UN Classifications Registry to include classification systems being used by other international and large-scale epistemic communities, as well as relevant national systems, thereby making available common standards and registries for all entities looking to make their data interoperable.	UNSC	By March 2019
	11	The newly-created GPSDD working group on citizen-generated data (CGD) should look to establish an inter-agency and expert group on CGD (or a "City Group") that can help to set standards and common methods for CGD to encourage greater uptake of CGD by NSOs. The group should promote the creation of CDOs within government who can help ensure a steady stream of high-quality CGD is being fed into the national data collection process.	GPSDD	By September 2017
Capacity & resources - Related to theory of change priorities 2 (closing persistent data gaps) and 5 (Closing the digital divide)	12	Establish a partnership between UNSC and UNESCO / GPE for data literacy training in schools.	UNSD	By March 2018
	13	Multilateral institutions, governments, and philanthropic donors should establish a global financing facility for statistics, which consolidates and focuses the inflow of data-related resources to capacity constrained countries and statistical systems and the production of global standards, as well as leverages private investment for SDG monitoring.	GPSDD to instigate dialogue. Lead financial management institution to be identified.	Initiated by early 2019
	14	GPSDD-facilitates a global public-private sponsorship platform for national statistical capacity.	GPSDD	Launched at WEF, January 2018

Annex 1: Using academic inputs to model estimates and help to fill data gaps

In instances where data is deficient, or does not provide insights for policy makers, model-based estimation provided by academics and experts can be a useful interim tool. However, it should not be used without full understanding of its limitations.

First, the results from modeling exercises are the output from models of widely varying complexity and sophistication. Models predicated on overly simplistic assumptions, or a failure to adequately represent the underlying processes, may produce structurally biased results. In recent years, the move to Bayesian and Monte-Carlo statistical approaches has improved our understanding of the uncertainty inherent in modeling exercises.[75] Yet these uncertainty bounds are often used to bolster claims about the robustness and accuracy of the modeling exercise. What is seldom, if ever, taken into account in these measures of uncertainty is that the actual structure of the model itself may be wrong.

Second, there is a real risk that, in producing model indicators for every country at a national or subnational level, there may be insufficient expertise to assess the results in light of known national features or to tweak the inputs used. This can result in substantial model error. The Global Burden of Disease (GBD) project run by the University of Washington's Institute for Health Metrics and Evaluation has substantial expertise and a commensurate international reputation in the modeling of health outcomes and metrics. Along with others, they have produced estimates of child mortality for developing countries across the globe. Yet these estimates are model-based, using variables such as gross domestic product (GDP) per capita as inputs. To take a single country as an example: South Africa has reasonably reliable estimates of GDP per capita. However, this metric masks a highly unequal distribution of access to resources, wealth and income across the country. Consequently, the estimates of child mortality do not take into account the strongly distinct intersections between poverty, disease and HIV (and associated high rates of child mortality) within heterogeneous groups within the country, resulting in estimates of child mortality that may be significantly at odds in both level and trend with estimates derived by epidemiologists, public health professionals and demographers within the country.

A third, more abstract risk associated with reliance on modeling exercises to monitor or track the SDGs and SDG indicators is that the process of producing those estimates may disempower systematically the process of data production in the Global South. There is a real danger that elite consortia from the Global North will increasingly tell the countries of the South what their progress towards meeting the SDGs is in the absence of empirical, locally-produced data. Partnerships between the North and South to share and spread knowledge and expertise is essential, as is model transparency and validation of all data used as input.

Finally, the scientific requirement of reproducibility means that data inputs, models, and data outputs used in the monitoring of SDG progress, should not be 'black boxes' but should be open-source and in the public domain.

Despite these concerns, it is equally clear that there is a need to use model-based indicators to enable immediate monitoring and tracking of the SDGs. Since the bulk of the expertise for producing these indicators lies not only in the North, but also outside of the formal national and supranational statistical system and framework, engagement with these groupings can and should be used as a case study for improving collaboration across actors and sectors.

Source: Written by Tom Moultrie, IUSSP and Virginia Murray, Public Health England

References

1. Harris, P. (2017). Data – the thing that ties it all together. 22 May, Committee for Economic Development of Australia (CEDA), Melbourne.

2. United Nations General Assembly resolution 70/1, Transforming our world: the 2030 Agenda for Sustainable Development, A/RES/70/1 (21 October 2015). Available at: undocs.org/A/RES/70/1.

3. Cassidy, M. (2014). Towards assessing gaps in indicator availability and coverage. [online] New York: Sustainable Development Solutions Network (SDSN). Available at: http://unsdsn.org/wp-content/uploads/2014/07/Assessing-Gaps-in-Indicator-Availability-and-Coverage.pdf [Accessed 31 Aug. 2017].

4. Espey, J., et al., (2015). Data for development: a needs assessment for SDG monitoring and statistical capacity development. New York: Sustainable Development Solutions Network. Available at: http://unsdsn.org/resources/publications/a-needs-assessment-for-sdg-monitoring-and-statistical-capacity-development/ [Accessed 5 June 2017]

5. There have been valiant efforts to map out a future data ecosystem and to signpost the necessary direction of travel. In November 2014 the Independent Expert Advisory Group on a Data Revolution for Sustainable Development outlined many of the emerging challenges and set out a series of principles to guide government and third party efforts to achieve a data revolution in their report, A World That Counts. In 2015 a consortium of data actors, led by SDSN and Open Data Watch, produced the first comprehensive estimate of what it would cost to monitor the SDGs in 77 IDA-recipient countries. Recently, a High Level Group of the UN Statistical Commission presented a Global Action Plan in Cape Town, South Africa, to reform national statistical systems to achieve the 2030 Agenda. The Global Action Plan was adopted by the UN Statistical Commission in March 2017 and will be presented to ECOSOC in July 2017. It outlines six areas for improvement: strategic leadership, modernization of national statistical systems, strengthening basic statistical activities, dissemination of data, new partnerships for data production and use, and the mobilization of resources.

6. See references above.

7. United Nations. (2016.) Report of the Inter-Agency and Expert Group on Sustainable Development Goal Indicators. Report No.: E/CN.3/2016/2/Rev.1. [online] New York: United Nations Statistics Division. Available from: http://unstats.un.org/unsd/statcom/47th-session/documents/2016-2-IAEG-SDGs- Rev1-E.pdf [Accessed 31 Aug. 2017].

8. SDG Indicators 2017, United Nations. Available from: https://unstats.un.org/sdgs/indicators/indicators-list/ [Accessed 31 Aug. 2017].

9. https://unstats.un.org/unsd/publication/SeriesM/Seriesm_67rev2e.pdf

10. United Nations. (2017). Population and vital statistic report. Report No.: ST/ESA/STAT/SER.A/263. [online] New York: United Nations. Available at: https://unstats.un.org/unsd/demographic/products/vitstats/Sets/Series_A_2017.pdf [Accessed 31 Aug. 2017].

11. See for example World Health Organization. (2014) Improving Mortality Statistics Through Civil Registration and Vital Statistics Systems, Outcomes of a Technical Meeting, Geneva 2014. Available at: http://www.who.int/healthinfo/civil_registration/CRVS_MortalityStats_Guidance_Nov2014.pdf [Accessed 31 Aug. 2017]

12. Sustainable Development Solutions Network (SDSN) 2017, Target 6.1 by 2030, achieve universal and equitable access to safe and affordable drinking water for all, <http://indicators.report/targets/6-1/> [Accessed 31 Aug. 2017].

13. Wescoat, J. L., S. Fletcher, and M. Novellino. (2016). National Rural Drinking Water Monitoring: Progress and Challenges with India's IMIS Database. In: Water policy, vol. 18, no. 4. [online] United Kingdom: IWA Publishing. p. 1–18. Available at: http://wp.iwaponline.com/cgi/doi/10.2166/wp.2016.158 [Accessed 31 Aug. 2017].

14. World Health Organization (WHO) and UNICEF. (2017). Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines. [online] Geneva: World Health Organization (WHO) and UNICEF. License: CC BY-NC-SA 3.0 IGO. Available at: http://apps.who.int/iris/bitstream/10665/258617/1/9789241512893-eng.pdf?ua=1 [Accessed 31 Aug. 2017].

15. World Health Organization (WHO) and UNICEF. (2017). Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines. [online] Geneva: World Health Organization (WHO) and UNICEF. License: CC BY-NC-SA 3.0 IGO. Available at: http://apps.who.int/iris/bitstream/10665/258617/1/9789241512893-eng.pdf?ua=1 [Accessed 31 Aug. 2017].

16. See, for example: United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG). (2014). A world that counts. [online] Independent Expert Advisory Group Secretariat.

Available at: http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf [Accessed 31 Aug. 2017].; Sustainable Development Solutions Network (SDSN) et al. (2015). Data for development. New York: Sustainable Development Solutions Network. Available at: http://unsdsn.org/wp-content/uploads/2015/04/Data-for-Development-Full-Report.pdf [Accessed 31 Aug. 2017].; Center for Global Development. (2014). Delivering on the data revolution in sub-Saharan Africa. [online] Washington, D.C.: Center for Global Development. Available at: https://www.cgdev.org/sites/default/files/CGD14-01%20complete%20for%20web%200710.pdf [Accessed 31 Aug. 2017].; Asian Development Bank (ADB). (2010). Administrative data sources for compiling Millennium Development Goals and related indicators. [online] Manila: Asian Development Bank. Available at: https://www.adb.org/sites/default/files/publication/28297/sources-compiling-mdg.pdf [Accessed 31 Aug. 2017]..

17. Jack Levis: The hardest step in innovation? Looking foolish in front of the crowd, TED Talk, available at: https://www.ted.com/watch/ted-institute/ted-ups/jack-levis-the-hardest-step-in-innovation [Accessed 31 Aug. 2017]

18. United Nations General Assembly resolution 70/1, Transforming our world: the 2030 Agenda for Sustainable Development, A/RES/70/1 (21 October 2015). Available at: undocs.org/A/RES/70/1.

19. "Planetary boundaries define the safe operating space for humanity in the Earth system along nine critical dimensions: greenhouse gas emissions, nitrogen and phosphorus loading, ozone depletion, chemical pollution, freshwater use, ocean acidification, land use change, aerosol loading, and loss of biodiversity." SDSN (2014) An Action Agenda for Sustainable Development: Report for the UN Secretary-General, Sustainable Development Solutions Network: New York, pp.2. See also Steffan, W. et al., (2015) Planetary boundaries: Guiding human development on a changing planet, Science, Vol. 347, Issue 6223. Available at: http://science.sciencemag.org/content/347/6223/1259855 [Accessed DD Mon. YYYY]

20. The medium population forecast of the UN Population Division assumes a substantial drop in fertility rates. If this drop does not occur then population might increase beyond 9 billion by 2050.

21. United Nations Office for Disaster Risk Reduction (UNISDR). Sendai framework for disaster risk reduction 2015-2030 (18 March 2015). Available at: http://www.unisdr.org/files/43291_sendaiframeworkfordrren.pdf.

22. Murray, V. et al. (2017). Coherence between the Sendai Framework, the SDGs, the Climate Agreement, New Urban Agenda and World Humanitarian Summit and the role of science in their implementation. [online] International Council for Science (ICSU) and Integrated Research on Disaster Risks (IRDR). Available at: https://www.icsu.org/cms/2017/05/DRR-policy-brief-5-coherence.pdf [Accessed 31 Aug. 2017].

23. See: FutureLearn, Predicting the spread of epidemics, accessed 31 Aug. 2017, <https://www.futurelearn.com/courses/big-data/0/steps/7642>. However, it should be noted that Google is no longer releasing its flu prediction data on account of a series of challenges with the data. See also: United Nations Foundation. (2015). Data disaggregation: like the layers of a pyramid. [Blog] UN Foundation Blog - Global Connections. Available at: http://unfoundationblog.org/data-disaggregation-like-the-layers-of-a-pyramid/ [Accessed 31 Aug. 2017].

24. Sustainable Development Solutions Network (SDSN) 2017, Target 3.7 by 2030 ensure universal access to sexual and reproductive health care services, including for family planning, information and education, and the integration of reproductive health into national strategies and programs, accessed 31 Aug. 2017 , <http://indicators.report/targets/3-7/>

25. CNN (2017) Facebook is closing in on the 2 billion user mark. Blog post and video, available at: http://money.cnn.com/2017/02/01/technology/facebook-earnings/index.html

26. United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG). (2014). A world that counts. [online] Independent Expert Advisory Group Secretariat. Available at: http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf [Accessed 31 Aug. 2017]

27. Cape Town Global Action Plan for Sustainable Development Data, United Nations Statistics Division, accessed 31 Aug. 2017, <https://unstats.un.org/sdgs/hlg/Cape-Town-Global-Action-Plan/>

28. Prydz, E. (2013). 'Knowing in time: how technology innovations in statistical data collection can make a difference in development' in Discussion paper for session 3.2 - 2013 Organization for Economic Cooperation and Development (OECD) Global Forum on Development. Available at: www.oecd.org/site/oecdgfd/. [Accessed 31 Aug. 2017].

29. Laskowski, N. (2015). 'Will the role of the chief data officer eclipse that of the CIO?' SearchCIO, July. Available at: http://searchcio.techtarget.com/opinion/Will-the-role-of-the-chief-data-officer-eclipse-the-CIO [Accessed 31 Aug. 2017].

30. Open Data Watch. (2017). Open data to support the Sustainable Development Goals. Open Data Watch. Available at: http://opendatawatch.com/wp-content/uploads/2017/03/UNSC2017-OpenDataSupportsSDGs.pdf [Accessed 31 Aug. 2017].

31. As above.

32. Sustainable Development Solutions Network (SDSN) 2017, Target 17.18 by 2020, enhance capacity building support to developing countries, including for LDCs and SIDS, to increase significantly the availability of high-quality, timely and reliable data disaggregated by income, gender, age, race, ethnicity, migratory status, disability, geographic location and other characteristics relevant in national contexts, accessed 31 Aug. 2017, <http://indicators.report/targets/17-18/>

33. Sustainable Development Solutions Network (SDSN) 2017, Target 17.19 by 2030, build on existing initiatives to develop measurements of progress on sustainable development that complement GDP, and support statistical capacity building in developing countries, accessed 31 Aug. 2017, <http://indicators.report/targets/17-19/>

34. Voluntary National Reviews, United Nations, accessed 31 Aug. 2017, <https://sustainabledevelopment.un.org/vnrs/>

35. http://www.statsghana.gov.gh/docfiles/SDGs/Data%20Roadmap%20Forum-%20Concept%20Note.pdf

36. Ghana Statistical Service and the Global Partnership for Sustainable Development Data (GPSDD). (2017). Ghana's Data for Sustainable Development Roadmap Forum, April 15-17th, 2017, Concept Note. Available at: http://www.statsghana.gov.gh/docfiles/SDGs/Data%20Roadmap%20Forum-%20Concept%20Note.pdf [Accessed 31 Aug. 2017]

37. For more information, see: Global Partnership for Sustainable Development Data (GPSDD) 2016, OPAL, accessed 31 Aug. 2017, <http://www.data4sdgs.org/dc-opal/>

38. United Nations Statistical Commission (2017). Participation of civil society, Event Page. Available at: https://unstats.un.org/unsd/statcom/48th-session/resources/cso-guide/ [Accessed 31 Aug. 2017]

39. United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG). (2014). A world that counts. [online] Independent Expert Advisory Group Secretariat. p. 3. Available at: http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf [Accessed 31 Aug. 2017]

40. Cape Town Global Action Plan for Sustainable Development Data, United Nations Statistics Division, accessed 31 Aug. 2017, <https://unstats.un.org/sdgs/hlg/Cape-Town-Global-Action-Plan/>

41. For more information on City Groups, visit: UN Statistical Commission, 2017, City Groups, accessed 20 Aug. 2017, <https://unstats.un.org/unsd/methods/citygroup/>

42. Banisar, D. National Comprehensive Data Protection/Privacy Laws and Bills 2016 (November 28, 2016). Available at SSRN: https://ssrn.com/abstract=1951416 or http://dx.doi.org/10.2139/ssrn.1951416

43. Talbot, D. (2014) Cell-phone Data Might help Predict Ebola's Spread, Rewriting Life, MIT Technology Review. Available at: https://www.technologyreview.com/s/530296/cell-phone-data-might-help-predict-ebolas-spread/ [Accessed 29 Aug. 2017]

44. For more information visit: https://www.ons.gov.uk/aboutus/whatwedo/paidservices/virtualmicrodatalaboratoryvml and https://www.gesis.org/gml/european-microdata/ [Last Accessed 1 Sept. 2017]

45. A recent report by the Center for Open Data Enterprise (ODE) reported that NSOs frequently do not trust private companies producing 'big data,' as they do not believe it produces methodologically sound results. The Center for Open Data Enterprise (ODE). (2016). Briefing paper on open data for public-private collaboration. Available at: http://reports.opendataenterprise.org/BriefingPaperonOpenDataforPublic-PrivateCollaboration.pdf [Accessed 5 June 2017]

46. Sinai, N. and Dole, A. (2014). 'Leading pharmacies and retailers join Blue Button initiative' HealthIT Buzz, Feb. 7. Available at: http://www.healthit.gov/buzzblog/consumer/leading-pharmacies-retailers-join-blue-button-initiative [Accessed 17 May 2016]; featured in: The Center for Open Data Enterprise (ODE). (2016). Briefing paper on open data for public-private collaboration. Available at: http://reports.opendataenterprise.org/BriefingPaperonOpenDataforPublic-PrivateCollaboration.pdf [Accessed 5 June 2017]

47. Partnership for Resilience & Preparedness (PREP) 2017, PREP, accessed 31 Aug. 2017, <http://www.prepdata.org/>

48. Orrell, T. (2017). What are the principles of joined-up data? [Blog] Joined-up Data Standards. Available at: http://juds.joinedupdata.org/ [Accessed June 5, 2017]

49. Wilson, C. & Rahman, D. (2015). Citizen generated data and governments: towards a collaborative model, Policy Brief, Civicus - Datashift, accessed 31 Aug. 2017, <http://civicus.org/images/citizen-generated%20data%20and%20governments.pdf>

50. UN Global Pulse, (2015). Mining citizen feedback data for enhanced local government decision-making. Indonesia: UN Global Pulse. Available at: http://www.unglobalpulse.org/sites/default/files/UNGP_ProjectSeries_Citizen_Feedback_2015.pdf [Accessed 31 Aug. 2017].

51. For more information, visit: Dar Ramani Hura 2016, Dar Ramani Hura, accessed 21 Aug. <http://ramanihuria.org>

52. Partnership in Statistics for Development in the 21st Century (PARIS21). How to build statistical capacity. PARIS21. Available at: https://www.paris21.org/sites/default/files/Capacity-buiding-knowledge-series.pdf [Accessed 31 Aug. 2017].

53. For more information visit: www.paris21.org [Accessed 1 Sept. 2017]

54. Partnership in Statistics for Development in the 21st Century (PARIS21). How to build statistical capacity. PARIS21. Available at: https://www.paris21.org/sites/default/files/Capacity-buiding-knowledge-series.pdf [Accessed 31 Aug. 2017].

55. For more information, visit: https://wb-gpsdd-datainnovation.forms.fm/call-for-proposals-collaborative-data-innovations-for-sustainable-development [Accessed 31 Aug. 2017].

56. See reference above.

57. Data Pop Alliance. (2015) Beyond data literacy: reinventing community engagement and empowerment in the age of data. Available at: http://datapopalliance.org/item/beyond-data-literacy-reinventing-community-engagement-and-empowerment-in-the-age-of-data/ [Accessed 31 Aug. 2017].

58. Internet Live Stats (www.InternetLiveStats.com) Elaboration of data by International Telecommunication Union (ITU), World Bank, and United Nations Population Division. [Accessed 29 Aug. 2017

59. The United States faces a shortage of between 140,000 and 190,000 professionals with analytical expertise, and 1.5 million managers and analysts who can make decisions based on big data analysis, according to research by McKinsey, the management consultancy. As featured in the Washington Post: Ravindranath, M. (2014) Elementary schools start teaching data literacy, The Washington Post, Nov. 16. Available at: https://www.washingtonpost.com/business/on-it/elementary-schools-start-teaching-data-literacy/2014/11/14/9e8bb4ec-6aac-11e4-9fb4-a622dae742a2_story.html?utm_term=.6ac047803f0b [Accessed 31 Aug. 2017]

60. http://www.cgdev.org/sites/default/files/CGD14-01%20complete%20for%20web%200710.pdf

61. Center for Global Development. (2014). Delivering on the data revolution in sub-Saharan Africa. [online] Washington, D.C.: Center for Global Development. Available at: https://www.cgdev.org/sites/default/files/CGD14-01%20complete%20for%20web%200710.pdf [Accessed 31 Aug. 2017]

62. Jerven, M. (2013). Poor numbers: How we are misled by African development statistics and what to do about it. Ithaca: Cornell University Press.

63. Greenwell, G. et al., (2016).'Making data portals work for the SDGs: A view on deployment, design and technology' in Discussion paper no. 8 – Partnership in Statistics for Development in the 21st Century (PARIS21). Available at: https://www.paris21.org/sites/default/files/Paper_on_Data_Portals%20wcover_WEB.pdf [Accessed 31 Aug. 2017]

64. Global Partnership for Sustainable Development Data (GPSDD). (2016). The state of development data funding 2016. Available at: http://opendatawatch.com/the-state-of-development-data-2016/ [Accessed 31 Aug. 2017]

65. See reference above.

66. See reference above.

67. See reference above.

68. PRESS2016, Partnership in Statistics for Development in the 21st Century (PARIS21), accessed 31 Aug. 2017, <http://www.paris21.org/PRESS2016>

69. See reference above.

70. Global Partnership for Sustainable Development Data (GPSDD). (2016). The state of development data funding 2016. Available at: http://opendatawatch.com/the-state-of-development-data-2016/ [Accessed 31 Aug. 2017]

71. See reference above.

72. See, for example: Espey, J., et al., (2015). Data for development: a needs assessment for SDG monitoring and statistical capacity development. New York: Sustainable Development Solutions Network. Available at: http://unsdsn.org/resources/publications/a-needs-assessment-for-sdg-monitoring-and-statistical-capacity-development/ [Accessed 5 June 2017]

73. Sachs, J. & Schmidt-Traub, G. (2017). 'Global Fund lessons for Sustainable Development Goals,' Science, vol. 356, no. 6333. p. 32-33. Available at: http://science.sciencemag.org/content/356/6333/32?ijkey=TVyxfcW7Uco9M&keytype=ref&siteid=sci [Accessed 5 June 2017]

74. See references above.

75. A helpful summary of these methods is provided by UNSW's School of Mathematics and Statistics. Available at: https://www.maths.unsw.edu.au/research/bayesian-and-monte-carlo-methods [Accessed 29 Aug. 2017].

SUSTAINABLE DEVELOPMENT SOLUTIONS NETWORK
A GLOBAL INITIATIVE FOR THE UNITED NATIONS

TRENDS
Thematic Research Network on Data and Statistics

giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

WILLIAM + FLORA Hewlett Foundation