UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT

DECEMBER 2017

UNCTAD Research Paper No. 13
UNCTAD/SER.RP/2017/13/Rev.1

Statistical Capacity Building for Sustainable Development: System prerequisites

Steve MacFeely & Nour Barnat

Statistics and Information Branch, Division on Globalization and Development Strategies, UNCTAD

steve.macfeely@unctad.org

Abstract

This paper argues that to prioritise the data requirements of the Sustainable Development Goal (SDG) monitoring framework, requiring 230 global indicators and spanning the full spectrum of development issues, over the development of national statistical systems may be a mistake. Rather, countries should prioritise the development of efficient national statistical systems that are sufficiently flexible, responsive and affordable to satisfy the enormous appetite of the SDG monitoring framework but also national and regional information requirements. The growing recognition of the importance of good quality, independent official statistics to support development and progress, provides a unique opportunity to make a real and long-lasting investment to improve national statistical systems. This will require coordinated investment and political support from donors and international organisations. The prerequisities for a modern statistical system are outlined. Without these countries will not be able to build statistical systems appropriate to a data driven world nor meet existing and future demands for information, including the SDG monitoring framework.

Key words: Legal framework, Institutional environment, Data infrastructure, SDGs

Contents

Acknowledgements

Comments from John Dunne (CSO Ireland), Kieran Walsh (ILO) and Johannes Jütting (Paris21) for feedback and comments.

Introduction

In 2015, the United Nations (UN) Millennium Development Goals (MDGs) came to the end of their 15 year lifecycle and were replaced by the considerably more ambitious UN Agenda 2030 for Sustainable Development and the Sustainable Development Goals (SDGs). From a statistical perspective the implications of Agenda 2030 for the accompanying monitoring framework are enormous, as not only have the number of goals and targets increased considerably (The MDGs had 8 Goals, 21 targets and 60 indicators whereas the SDGs have 17 Goals and 169 targets and 230 indicators) but so also has the complexity of these targets. The scope of Agenda 2030 is also far wider than their predecessor, attempting to span the full spectrum of development issues, including not only aspects of society, economy and the environment but also institutional coordination. This massive increase in scope and scale raises real questions regarding the capacity of national and international statistical systems or what others have described as 'data ecosystems' (Jütting, 2016) to implement such an enormous monitoring framework. The complexity and ambition of this challenge led Mogens Lykketoft, President of the UN General Assembly to describe it as an *'unprecedented statistical challenge'* (Lebada, 2016).

Costs and benefits

The sheer scale of the SDG monitoring framework raises questions of cost-benefit. For example, the Sustainable Development Solutions Network estimate that $1 billion per annum *'will be required to enable 77 of the world's lower-income countries to catch-up and put in place statistical systems capable of supporting and measuring the SDGs'* (Sustainable Development Solutions Network, 2015: 10). Paris 21 (2015: 11) estimated that *'funding for statistics needs to be increased from current commitments of between US$300 million and 500 million to between US$1 billion and 1.25 billion by 2020'.* Danish economist Bjorn Lomborg estimates that the SDG monitoring framework could cost as much as $1.5 billion per target, meaning that measurement would account for 12.5% of total development aid (The Economist, 2015). While clearly the cost estimates vary, there is some consensus that populating the SDG monitoring framework will be expensive.

The danger of such a costly exercise is that it may divert scare statistical resources from where they might be elsewhere used and perhaps with more relevance at a national level. The result is that the development of other nationally important statistics might be retarded or stopped altogether. The OECD (2015: 20) put it well *'A concerted effort from the international community over the next 15 years will be needed to ensure that SDG monitoring does not impose inordinate costs on developing countries or divert resources from achieving national statistical development strategies'*. This prompts the question as to whether it is worth populating SDG indicators at the expense of others. And this is a pertinent question, as many of the SDG indicators have proven unsatisfactory or even controversial. In part, this dissatisfaction arises as several of the SDG targets themselves are very complex and would in truth require several indicators to do them justice. Other targets are ambiguous leaving many unclear as to what precisely is meant. So how much resource should be dedicated to fulfilling SDG data requirements as opposed to developing economically sustainable and efficient national statistical systems (NSSs)? It is a difficult question to answer as the two objectives are not necessarily mutually exclusive. But developing a NSS will potentially address at least some of the SDG requirements by default. It is not clear that the converse is true however. For countries that do not yet have a NSS, too much emphasis on the SDG indicators, however important or well intentioned, may be a distraction and ultimately counter-productive.

Furthermore, it is important to understand that the 230 SDG indicators are only performance metrics and the data required to populate them are only a fraction of the data required to implement the SDGs. Policy formulation requires more complex and nuanced data that allows several factors to weighed-up before

informed decisions can be made. So we must draw a distinction between the output data required to populate selective performance metrics and the input data required to formulate policy that will help to implement the SDGs. A balance must be found whereby NSSs can be developed in parallel with the systems required to generate both the input data required to support policy formulation and the output data required to populate the SDG indicators (or at least a prioritised set for individual countries).

Change is constant

The challenge is how to weigh the demands of Agenda 2030 with the multitude of other national data priorities that exist. As countries progress along the development spectrum or find themselves on different slopes of the business cycle, their data and information requirements, will be different and will change. Agenda 2030, like the MDGs before it, has a 15 year lifespan. In the context of history, 15 years is the blink of an eye. Nevertheless we can safely predict that, for many individual countries in various parts of the world, during the next 15 years, many unanticipated events will unfold and some of these will have a dramatic and profound impact on national priorities and development plans. Murphy's Law applies: *'what can happen, will happen'*. Natural disasters, such as, tsunami, earthquakes and famines; Political or economic events, such as, wars, the unravelling or creation of political unions, the emergence of new countries or new financial or economic crises may all play their part in shifting international or national priorities. The sudden migration crisis in Europe triggered by the conflict in Syria is a perfect example. Thus we can reasonably anticipate that national and global priorities will differ and change over the next 15 years. The challenge for countries, and for the international community, is how to build affordable statistical systems that are sufficiently flexible and responsive to these evolving priorities but can at the same time satisfy the enormous appetite of the SDG monitoring framework. This poses a dilemma - in designing statistical capacity building programmes, how can a global indicator and monitoring framework (with urgent data demands) be designed so that it does not suffocate or overwhelm national and regional priorities (with current and long term data needs) but rather feeds off them? While pondering this dilemma, it is worth recalling that capacity building is most effective when it is home-grown, long-term in perspective and managed collectively by those who stand to benefit (Jütting**,** 2016). It is also worth considering the impact on international organisations, many of whom will be faced with the problem of how to avoid destabilising existing frameworks and programmes that have taken years to fund and develop.

Interlinkages and interdependencies

Another important discriminating factor between the ambitions of the SDGs compared with that of the MDGs is the desire to reflect the interconnectedness of the social, economic, environmental and institutional pillars. Thus the SDG monitoring framework must be integrated in a way that is capable of illustrating and highlighting interlinkages across statistical domains. This raises another question - how can countries be helped to develop NSSs that are capable of measuring and explaining inter-linkages or interdependencies between the various strands of Agenda 2030 and between their own national and regional activities? While it is tempting to focus capacity building programmes and resources towards specific targets, such a narrow and short-term approach will not address this challenge and will come at the cost of developing wider and more flexible statistical systems. We argue that the balance should be tipped in favour of developing of strong, sustainable NSSs. Statistical systems take time to develop and mature but putting in place strong NSSs will yield more long lasting results for developing countries. Investing in a sound foundation, albeit generic in scope, will ultimately contribute more to the availability of information in the longer term than focusing on a set of specific contemporary targets. A properly functioning system will be better able to contribute to all statistics and not just the specific set chosen for the SDG monitoring framework.

Organisation of the paper

The remainder of this paper is presented in six sections. Section one provides definitions of a NSS and administrative data. It also briefly explains why each of the fundamental pillars is core to a NSS. The following three sections (2, 3 and 4) outline each of the pillars: legal framework; institutional environment; and national data infrastructure in turn. Section five makes the link between capacity building and NSSs and

in particular stresses the need for close coordination across international organisations. The sixth and final section concludes the paper.

1. National Statistical Systems

A NSS can be defined as the group of organisations or agencies (or specialist statistical units within those organisations or agencies) who together collect, process and disseminate official statistics on behalf of the government of a country. Or more comprehensively, organisations or agencies producing any of the steps identified in the Generic Statistical Business Process Model (United Nations Economic Commission for Europe, 2013) on behalf of the government of a country. The United Nations Statistics Division has long championed the importance of NSSs in their Handbook of Statistical Organisation (United Nations Statistics Division, 2003), the latest version being no exception. Paris21 (2004) too has advocated this cause, developing guidelines for NSSs back in 2004 - the 'National Strategies for the Development of Statistics'. More recently the Sustainable Development Solutions Network (2015) has stressed the importance of NSSs for the production of official statistics. So too has the Addis Ababa Action Agenda of the Third International Conference on Financing for Development in July 2015, noting *'National statistical systems have a central role in generating, disseminating and administering data'* (United Nations, 2015a: 58).

Statistical systems can be described as either centralised or decentralised depending on the extent to which responsibility for compiling official statistics rests with the central institution (usually the NSO). Statistical systems are said to be centralised when all, or most, of the products of the statistical system are compiled and disseminated by the central organisation. A decentralised system may have statistical units embedded in a number and variety of Government ministries and agencies spread across a public service. While there are strengths and weaknesses to both approaches, centralised statistical systems are typically independent and impartial but sometimes struggle with relevancy, owing to their remoteness from policy debate. In contrast, the reverse is often true of decentralised statistical units located in Government ministries, as their work is highly relevant but is more susceptible to political interference and pressure to present statistics relating to Ministerial policies and outcomes in a favourable light, thus compromising the credibility of the data (Paris 21, 2005).

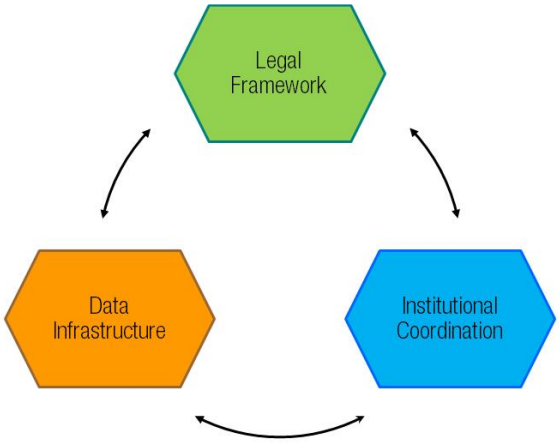
Before going any further, it is also useful to explain what we mean by administrative and secondary data. Administrative data or public sector data can be defined as data that are collected primarily for administrative or management purposes. National public administrations typically collect, maintain and update sizeable volumes of data as a matter of routine for the purposes of registration, transaction, and record keeping associated with the delivery of a service. These data pertain to the wide range of administrative functions in which states are involved, ranging from individual and enterprise tax payments to social welfare claims, education participation and grants, farming subsidies or individual health records. Typically these administrative records are collected and maintained at the lowest level of aggregation i.e. transaction or interactions by individual taxpayer/applicant/recipient with the state, making these data very rich from an analytical perspective. Secondary data is a broader concept that includes any data that have already been collected or available from other sources. Secondary data is not limited to administrative data and can also include non-government, commercial and research data. Secondary data may sometimes be cheaper and occasionally more quickly obtainable than primary data (assuming legal protocols and mechanisms to facilitate institutional cooperation are in place) and often will have very extensive population coverage.

1.1 Three fundamental pillars

While there are many diverse aspects required of a NSS, such as finance, governance and oversight, we argue there are three fundamental pillars that are absolutely core to a modern NSS (see Figure 1). First, and perhaps most importantly, a robust legal framework must exist. The compilation of official statistics must be governed by national legislation that ensures the professional independence of the statistical system and the head of that system in order to provide protection against political interference or politicisation. Legislation must also give the statistical system power to compel businesses and persons to provide information for statistical purposes. As a quid pro quo, the confidentiality of information provided under the auspices of statistical law must be protected. Finally a modern statistical system must have right of access to

administrative and secondary data sources - without which, the system will not be able to meet the growing demands of users.

Figure 1
Fundamental Pillars of a National Statistical System



Secondly, there must exist an environment where institutions of the statistical system can cooperate and coordinate their activities with other specialist and sub-national and international agencies and organisations. Sanga (2011), discussing the situation in African countries, identified lack of statistical coordination within NSSs and between NSSs and international organizations as one of the factors that hindered the capacity of countries to measure MDG performance. Most likely it will be identified as a retarding factor for the SDG process in years to come, unless addressed now. NSSs in the future will require technical expertise and information held by organisations outside the statistical system. Availing of new opportunities arising from new data sources or meeting the challenges of greater demands for climate related data, geo-spatial information, better visualisation or improved explanations will require new strategic partnerships, both nationally and internationally in order to get the mix of skills and experience necessary. As the volumes of data increase, owing to these new data sources, more thought and resources must be dedicated to the curation of data and this too may require external expertise, to ensure that comparable time series are preserved. International and supra-national organisations play an important role in collating official statistics on an internationally harmonised or comparable basis and facilitate the development of new concepts and production methods. It is vitally important that there are no impediments to NSSs and National Statistical Organisations (NSOs) cooperating fully with such organisations. Coordination between international organisations is also essential, to eliminate unnecessary duplication and to miminise the reporting burden imposed on NSSs, but this is outside the scope of this paper.

NSOs cannot possibly or efficiently deliver on the data demands facing them without building a supporting NSS. Landes (1999: 517), speaking about economics, reminded us that *'Economic analysis cherishes the illusion that one good reason should be enough, but the determinants of complex processes are invariably plural and inter-related. Mono causal explanations will not work'*. This is true of every subject, not just economics. To measure and illustrate the interconnectedness or interdependence of the various aspects of life, to identify multi casual explanations, requires that data from all different aspects of life can be connected or joined. In other words *'data must be comparable across time and sufficiently well described so as to facilitate integration with other data'* (Ribes and Jackson, 2013: 147), and to do this, data must be 'domesticated' i.e. the rituals and routines of collection and storage, of setting metadata standards must be established. Thus cooperation between specialist institutions and a common framework or infrastructure is absolutely necessary to facilitate data linking. If each compiler is proprietary of their data and will not share it (within obvious sensible limits and with appropriate legal and technical safeguards) the value of their data and everyone else's is greatly diminished. To realise the potential of data, data must be organised in a way that facilitates cross-pollination through the linking of microdata.

The final pillar for a functioning and efficient NSS is the existence of a National Data Infrastructure (NDI) that facilitates data-sharing and linking between the various agencies in that system. A NDI is essential for the efficient compilation of official statistics and to avoid what has been described as 'stranded data' (Singh, 2012). A NDI is probably the only realistic and affordable way of reaching the level of data granularity required to realise the ambition of 'no one gets left behind' (United Nations, 2015b) which has been translated for statisticians as 'leaving no one uncounted' (Lebada, 2016). But a NDI has much wider implications beyond improving the NSS. A NDI can also provide a platform to transform the way public administrations operate, facilitating not only improved public services and administrative efficiency but greatly improving the management information systems necessary to monitor and evaluate programmes. Central to the philosophy of an NDI is the ability to re-use and link public sector information. This is essential not only for the compilation of affordable statistics, but also the development of new dynamic indicators, small area and regional statistics. SDG Target 9.1, although presumably not drafted with data in mind, summarises the ambition of a NDI perfectly - *'Develop quality, reliable, sustainable and resilient infrastructure, including regional and trans-border infrastructure, to support economic development and human well-being, with a focus on affordable and equitable access for all*' (United Nations, 2015b: 20). Essential to the implementation of a NDI is to cultivate an appreciation or recognition of data as an asset and an essential part of the infrastructure necessary to support a modern public administration.

2. Pillar 1 - Legal Framework

The first pillar required to support a modern NSS is a sound legal framework. The implementation guidelines for the United Nations Fundamental Principles of Official Statistics states (United Nations Statistics Division, 2015: 62) *'High quality legislation is critical to the effective performance of a national statistical system*'. Irrespective of whether a statistical system is highly centralised or decentralised, all agencies or institutions belonging to that system, should have their activities supported by clear national legislation.

Activities like collecting private personal information, commercially sensitive business data or accessing confidential administrative or secondary microdata cannot be left to the vagaries of personal understandings or ad-hoc solutions. Equally, there can be no ambiguity regarding the safeguarding of confidentiality or the professional independence and impartiality of official statistics. Guarantees regarding such activities must be enshrined in national law so that all stakeholders understand clearly their rights and their obligations vis-a-vis the statistical system.

2.1 Statistical legislation

Statistical legislation will typically set out the roles and responsibilities of the key actors in the statistical system. For example, it should identify the head of the statistical system and define clearly that role. This will be particularly important if the head of the statistical system is different from the head of the NSO. NSSs operate within a wider national administrative and legal framework. So for a statistical system to properly function, the roles and relationships with other bodies, such as government departments or the central bank should be defined in legislation. Equally, the relationship and delineation between the statistical law and other legislative acts, such as, data protection and freedom of information should be clarified as together these legislative acts create an environment of public trust by protecting the information rights of individuals, upholding administrative transparency and permitting data to be collected and used for statistical purposes. Data protection legislation typically sets out law on the treatment, use and storage of personal data on identifiable living people. It is designed to protect individuals against possible misuse of information about them held by others. Freedom of Information legislation typically sets out law creating a full or partial public 'right of access' to previously unreleased information held by governments and public authorities. Thus statistics and public administration are to some extent intertwined and symbiotic, although statistical legislation has some clear and important distinctions and differences. These legislative acts should all be mutually supporting in so far as they should help to create an environment of public trust regarding the use of information.

While national statistical legislation should define and legislate for all aspects of a statistical system, there are, three absolutely necessary components: the professional independence of that system and in particular of the head of that system; The NSO and other statistical agencies of the NSS must be legally empowered to collect primary information for statistical purposes from persons, households, businesses and other institutions and have legal access to all appropriate administrative microdata held by national and regional public administrations for the purposes of compiling aggregate statistical information; and the safeguarding of confidentiality must be clearly set out in law.

2.2 Professional independence and impartiality

The professional independence and impartiality of a statistical system is vitally important to ensuring that the best data possible are compiled and disseminated without fear or favour. This is reflected in the first and second principles of the updated United Nations Fundamental Principles of Official Statistics (United Nations Economic and Social Council, 2013) which state:

> *Principle 1 - Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.*
>
> *Principle 2 - To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.*

and principle 6 of the revised European Statistics Code of Practice (Eurostat, 2011) which states:

> Principle 6 - *Statistical authorities develop, produce and disseminate European Statistics respecting scientific independence and in an objective, professional and transparent manner in which all users are treated equitably.*

It is essential that data providers trust the statistical system to only use data supplied for the purposes of compiling official statistics, and not for any form of legal or regulatory enforcement. Furthermore, it is also essential that official statistics are clearly independent of the political system and government interference or manipulation. Credibility will only come from independence (Citro and Straff, 2013). The United Nations Handbook of Statistical Organization is also uncompromising in this respect as *'It enjoins the chief statistician and staff to allow no interference by interested parties in the way facts are assembled and combined into statistics, nor in the method and timing of their release to the public'* (United Nations Statistics Division, 2003: 2). Independence must include secure and sustainable resourcing of the statistical system to avoid the