The Fundamental Pillars necessary for a Modern National Statistical Service

Steve MacFeely*
United Nations Conference on Trade and Development, Geneva, Switzerland
Centre for Policy Studies, University College Cork, Ireland
steve.macfeely@unctad.org

Abstract

In the excitement of the post-2015 Sustainable Development Goals, 'Big Data' and the 'Data Revolution' there is a danger that the statistical community is trying to run before everyone has learned to walk. The development of national and regional statistical systems is uneven and many basic challenges persist around the world. While it is proper, that developed countries forge ahead and take on the new challenges of big data and the more exotic aspects of the 'Data Revolution', it is important to remember than many NSIs cannot yet utilise basic data from their own national administrative systems to compile official statistics. Ironically, for many Least Developed Countries where information on progress towards the development goals is most needed the data are not available. This is not necessarily because of limitations in skills or even resources, but because some or all of what I have termed the 'fundamental pillars', required for a modern National Statistical System are not in place.

The data demands arising from the post-2015 Sustainable Development Goals are likely to be sizeable and complex. It is also likely that many national statistical institutes will struggle to deliver. In this paper I argue that national statistical institutes cannot possibly or efficiently meet these demands alone, and therefore it would be sensible for international organisations to actively promote and help develop coordinated national statistical systems or services. I argue there are three fundamental requirements or pillars necessary for a modern national statistical service, namely: a sound legal framework, a functioning and coordinated institutional environment and a national data infrastructure in support.

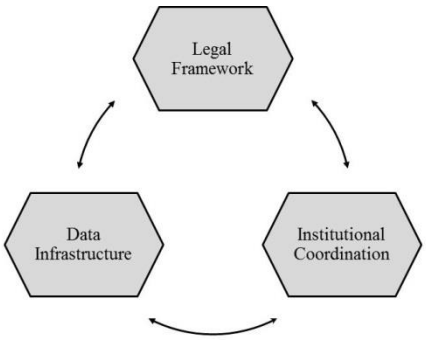
Keywords: Legal framework, institutional environment, data infrastructure

1. Introduction

Today in official statistics there is a significant push to modernise and transform the way data are being collected and statistics are compiled. Often these drives focus on the more exotic end of the modernisation spectrum, such as harnessing 'Big Data' or joining the 'Data Revolution' (IEAG, 2014). Both are sufficiently opaque in meaning to be universally attractive and unfortunately are viewed by many as cheap panaceas to the world's data gaps. While this 'gold rush fever' is understandable, I argue that it might be more prudent and ultimately more cost effective to first help countries to put in place the basic building blocks necessary to support modern statistical systems. Namely: statistical legislation; institutional coordination; and data infrastructure. This paper attempts to show the importance and inter-linkages of each of these key pillars (See Figure 1).

In this paper I argue that many national statistical institutes cannot possibly or efficiently build these foundations alone, and therefore it would be sensible for international organisations to actively promote and assist in the development of coordinated national statistical systems or services. While to some extent this is already done, through various domain specific programmes, there is no global programme that gives attention to the broad supports necessary i.e. there is no programme addressing the concept of a statistical system in the holistic sense.

Figure 1 - Fundamental Pillars of a National Statistical System



This paper is presented in four sections. The first three sections outline each of the pillars: legal framework; institutional environment; and national data infrastructure in turn. The fourth and final section concludes the paper.

2. Pillar 1 - Legal Framework

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 personal information, commercially sensitive business data or accessing sensitive administrative microdata cannot be left to the vagaries of personal understandings or ad-hoc solutions. Equally, guarantees regarding the safeguarding of confidentiality or professional independence or impartiality. These activities and guarantees must be enshrined in national law so that all stakeholders understand clearly their rights and their obligations vis-a-vis the statistical system.

Of course national statistical systems operate within a wider national administrative and legislative framework. The relationship of these governance frameworks and legislative acts to the statistical law plays an important, albeit indirect, role in the successful operation of a statistical service. For the purposes of this paper I will not address other legislation in detail, but limit my discussion to a narrower discussion on what should be included in primary statistical legislation. However the importance of 'Data Protection' and 'Freedom of Information' legislation to statistics should not be overlooked. Together these legislative acts create an environment of public trust by protecting the information rights of individuals while at the same time upholding administrative transparency and permitting data to be collected and used for statistical purposes. Thus statistics and public administration are to some extent intertwined and symbiotic, although statistical legislation has some clear differences, particularly regarding the flow of microdata.

While national statistical legislation should define and legislate for all aspects of a statistical system, there are I would suggest, three absolutely necessary components: Political and professional independence of that system and the Director-General or President of that system; The National Statistical Institute (NSI) and other statistical agencies of the system must be legally empowered to collect primary information for statistical purposes from persons, households, businesses and other institutions; and the NSI and statistical agencies of the system must also have legal access to all appropriate administrative microdata held by national and regional public administrations for the purposes of compiling statistical information. Furthermore, in order to future-proof any statistical legislation, I would argue that consideration must be given to allowing access to a broader concept of secondary data than has traditionally existed in many countries. Arguably, for the purposes of compiling official statistics, access to administrative data should now include access to secondary data i.e. secondary data would be defined to include not only administrative or public sector data but also

some important, commercially held data, such as 'big data' e.g. information on credit card transactions, information held by utilities or movements of mobile phones.

Finally, a functioning statistical system must protect the confidentiality of the persons and entities for which it holds individual data i.e. to guarantee the protection of identities and information supplied by all persons, enterprises or other entities. In effect this means that only aggregate data will be published by that statistical system and any aggregates will have been tested for primary and secondary disclosure. Furthermore, any legislation should also address the safe access to anonymised microdata for research purposes so that guaranteed confidentiality and anonmity are preserved. Although not usually part of the statistical legislation, most developed statistical systems publish supplementary security protocols (e.g. how paper or electronic questionnaires are destroyed after their useful shelf life has passed, or how records are stripped of unique official identifiers to safeguard protection) and procedures and in the interests of transparency. Such protocols are themselves frequently supplemented with public inventories of who has accessed microdata and for what purposes.

3. Pillar 2 - Institutional Environment

NSIs cannot reasonably be expected to meet the statistical and informational needs of a modern state in isolation. Increasingly, modern statistics are reliant on administrative and other secondary data. This is a trend that is likely to continue and escalate, as the costs of surveys become more prohibitive and the difficulties with hard to reach regions or cohorts intensify. The increasing demand for more complex multi-faceted analyses coupled with the growing appetite for dynamic rather than static data all point towards the importance of developing national statistical *systems*.

The importance of geography or space for properly understanding many social, environmental and economic phenomena is increasingly important. The marriage of statistics and space brings huge opportunities to analyse and disseminate statistics in new and interesting ways but it also requires careful coordination. Again, this is an important point as in some cases large administrative data may be the only way that good quality local or sub-national data can be compiled, even though sometimes this involves trade-offs between data availability and adherence to existing statistical concepts (MacFeely et al, 2013). This marriage also brings new challenges, particularly regarding the anonymisation of confidential cells, again highlighting the importance of sound legislation and protocols, particularly regarding the access to and use of microdata. The coordination of traditional statistical services with land and property registers, mapping and ordinance surveys and other geo-spatial services will be critical if the potential of geography and space is to be capitalised on from an information perspective, especially to yield good quality sub-national or local statistics.

What has been said about linking geography and statistics of course holds true for all sectors or domains. Landes (1998: p517), speaking about economics, reminds us *'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 for every domain. Thus a statistical system, if it is to meet properly the demands of its users must find ways to illustrate the interconnectedness or interdependence of various aspects of life. This is not easy, but it is certainly impossible without institutional cooperation.

At the national level, institutional coordination is essential. As noted above, a statistical system cannot properly and efficiently function without such coordination. NSIs or key compiling agencies must be legally allowed to access data held by the tax authorities in order to compile business registers and national accounts. Equally coordination between NSIs, Customs and Central Banks is essential if a good quality international trade and balance of payments statistics are considered desirable. But these

are just examples, coordination is required across all aspects of the administrative spectrum - education, health, energy, transport, environment, tourism and crime.

How institutional coordination is arranged will depend on the administrative structures, culture and legislation of a country. But a vitally important step is to formalise the coordination of roles and institutional responsibilities to provide clarity and ensure that cooperation is not dependent on particular individuals. Ideally, these arrangements should be supported by published MoUs so the exchange and use of data is transparent to everyone. Formal MoUs should also bring clarity by explicitly dealing with the storage, classification and access to micro-data. Several UN manuals offer some good advice in this regard - for example the UN International Merchanise Trade Statistics Manual (United Nations, 2011, p.59) states *'Effective institutional arrangments are usually characterised by (a) the designation of only one agency responsible for the dissemination of official […] statistics, (b) a clear definition of the rights and responsibilities of all agencies involved, and (c) the establishment of formalised working arrangements between agencies including agreements on holding inter-agency working meetings […] and on the access to micro-data that those agencies collect*'.

Coordination at regional level is every bit as important at national level as the institutional and legal relationships involved are often complex. To begin with the word 'region' may mean very different things depending on the context. From a statistical perspective, 'regional' typically means one of two things: sub-national or supra-national. But even this can result in a variety of different administrative arrangements. Sub-national coordination in a federal context may require very formal structures supported by additional legislation; for example, the Lande in Germany or the States in the USA. For other countries sub-national coordination may be less formal but no less complex and important for a variety of historical, cultural or economic reasons. Supra-national coordination is of course also very important, particularly for the compilation of comparable regional estimates, for political and economic unions like the Western African Economic and Monetary Union or the European Union.

The final pillar necessary within the institutional environment is international. Many of the issues discussed at national and particularly supra-national level, apply equally at the international level. The international statistical community plays an important role is helping to harmonise classifications, agree methodology and facilitate capacity building across countries. By collating data from around the world, they provide valuable data repositories and also help to identify implementation problems, for example, trade asymmetries. From a country perspective, it is very important to cooperate and coordinate activities with the international community. If countries act in an uncoordinated, unilateral way, it only serves to undermine the comparability and value of all the data, both national and international. The coordinated timing of taking censuses of population is a good example of how coordination is beneficial. Thus adopting internationally agreed calendars, standards and classifications is not only sensible but it also safeguards against manipulation of data and concepts for short-term national interests.

4. Pillar 3 - National Data Infrastructure

The final pillar, essential for a modern, national statistical service is a well organised and coherent National Data Infrastructure (NDI). Not only is such an infrastructure of paramount importance from a statistical perspective but it also critical for an efficient national administration (MacFeely & Dunne, 2014). As noted already, with so much attention being given to big data and data revolution, there is a risk that insufficient attention will be given to properly designing the basic architecture required to support national data or information systems. Without a properly designed foundation or infrastructure, many statistical capacity building programmes may not achieve their full potential or may simply be wasteful.

National public administrations typically collect, maintain and update sizeable volumes of data on a regular basis. 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 or education or farming grants. 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. The logic of a NDI is simply the organisation of this public or administrative data to maximise the potential value. Administrative data, if properly organised, are valuable assets and an essential part of the 'soft' infrastructure necessary to efficiently run a modern state and fuel a modern statistical service. Holders of other secondary data, such as private or commercial data should also be encouraged to adopt the same standards but the emphasis here is on public administrative data.

To develop a NDI a number of key databases must be developed. Specifically, three comprehensive databases or 'lists' are required: (i) a list of all persons in the state (with a unique ID attributed to each person), (ii) a list of all businesses in the state (each with a unique ID) and (iii) a list of all locations/buildings in the state (each with a unique ID and location co-ordinate). Furthermore, the inter-linkages between these lists are also required, so that the various interactions between them can be measured and understood e.g. where does a person live and work. Building a NDI on unique, permanent official and commonly used identifiers will permit public sector data to be analysed in a way that facilitates the identification of longitudinal, latitudinal, spatial and relational linkages. These linkages allow movements in time and space to be properly understood. Thus an 'object' or unit (e.g. individuals, enterprises or buildings) can be tracked over time as can their 'attributes' or characteristics (e.g. spatial location) and their relations to other units (e.g. family, employer, school, car). Thus the importance of a NDI, to both understanding geography and space and also for the development of dynamic indicators is clear. The importance of permanent or 'persistent' official identifiers is central to this approach.

For a comprehensive NDI to properly function, it must facilitate data-sharing and linking. The importance of being able to re-use and match public sector information cannot be overstated both for the compilation of modern official statistics but also for the efficient running of a modern state. Quite obviously, if the data made available to the NSI can be shared across the statistical system it will have a profoundly positive impact on the quality and range of official statistics that can be made available.

It is vital that the underlying data generated or associated with these services are organised in a coordinated way using the permanent public service identifiers and the same internationally agreed classifications and codes. By better organising and coordinating the management of administrative data, the potential of that information can be unlocked. To get maximum benefit from such an information system, the architectural design is crucial, and must involve the relevant permanent, official unique identifiers associated with each database or list. For those interacting with the state in any service or activity, use of these official identifiers must be mandatory. A move to such a universal design will de-silo existing systems. Only with such a system can the interactions and inter-relationships between citizens/business and the state be measured and understood.

5. Conclusions

This paper argues that it is absolutely essential to get the basic architectural design of national statistical systems right. Unfortunately not enough time and resources are being devoted to these fundamental or foundational issues as the majority of capacity building efforts are dedicated to domain specific capacity building programmes. The pressures created by the Data Revolution, Big Data and Sustainable Development Goals may exaggerate this situation as attention and resources may become further diverted and fragmented.

It is critical that in designing a national statistical system, the issues of legal framework, institutional environment and data infrastructure are considered. As far as is possible that design should consider the likely direction of Big Data and Data Revolution in order to future-proof that design. For example, statistical legislation must anticipate access beyond pure administrative data and should now consider including other secondary data, which includes Big Data. It is important to stress here, this paper does not argue against the use of Big Data, far from it, merely that the statistical system must have in place the three basic pillars so that access is legal and the data can be linked in order that the full potential of the data can be realised.

From an international agency perspective, arguably there is a need to agree on a more holistic and coordinated approach to statistical development and capacity building. Countries should be encouraged and assisted in implementing these three fundamental pillars. National governments should be helped to understand the connection between shared and linked information and 'seamless' or 'joined-up' government to improve not just official statistics but also public services and government performance by increasing administrative efficiency, transparency and improving policy formulation and assessment. The cost of not taking such a holistic approach to capacity building could be large and enduring.

References

Independent Expert Advisory Group on a Data Revolution for Sustainable Development (2014). 'A World That Counts: Mobilising the Data Revolution for Sustainable Development' Report prepared at the request of the United Nations Secretary-General.
Retrieved from: http://www.undatarevolution.org/report/

Landes, D. (1999). 'The Wealth and Poverty of Nations' Abacus, London.

MacFeely, S., J. Delaney & F. O'Donoghue (2013), 'Using business registers to conduct a regional analysis of enterprise demography and employment in the tourism industries: learning from the Irish experience', *Tourism Economics*, 2013, 19 (6), pp. 1293–1316.

MacFeely, S. & J. Dunne (2014), 'Joining up Public Service Information – The rationale for a National Data Infrastructure', *Administration*, Vol. 61, No. 4, pp. 93 – 107.