# Assessing the demand and supply of statistics in the developing world: some critical factors

Jeffery I Round

University of Warwick

May 2014

Abstract:

Discussions about the relationship between the demand and supply of statistics have taken place for many decades, especially in international statistical fora, but relatively little has been written that is in the public domain. To some extent this discussion, and hence this paper, is prompted by recent developments in the need to monitor international development goals. The recognition that evidence-based policy is overwhelmingly preferred strongly underpins an ever-increasing demand for official statistics and a need to enhance and prioritise the supply accordingly. There is a discussion about the concepts of supply of and demand for statistics in this context, underpinned by a recognition that official statistics have many of the characteristics of global public goods, which has implications for any assessment of demand for data. The supply side, conditioned by statistical capacity, has received most attention to date.

A paper prepared for the PARIS21 Secretariat[1].

[1] Acknowledgements: I wish to thank Johannes Jütting for suggesting I tackle this paper and also for his very helpful and challenging comments along the way; also, I thank Graham Eele for his perceptive comments on an earlier draft. I also thank two anonymous referees for their very helpful comments. However, I am entirely responsible for the content.

## 1. Introduction

Discussions about the relationship between the demand and supply of statistics have taken place for many decades, especially in international statistical fora, but relatively little has been written that is in the public domain. To some extent this discussion, and hence this paper, is prompted by recent developments in the need to monitor international development goals. The recognition that evidence-based policy is overwhelmingly preferred to any other kind of policy (Scott, 2005) strongly underpins an ever-increasing demand for official statistics and a need to enhance and prioritise the supply accordingly. Of course, assessing the relationship between supply and demand does not fully account for other issues, for example governance and institutional issues, but it does provide an initial framework for analysis.

In section 2 there is a discussion of the basic concepts underlying the demand and supply of statistics to help us to understand the fundamentals. We pursue the implications of the argument that most official statistics have all the characteristics of global public goods and that this is crucial in understanding the key factors influencing their demand and supply and how to match them in practice. Non-official statistics are different; they are more likely to be marketed and the supply and demand for them are influenced by market conditions. Trends in the supply and demand for official statistics over time are briefly examined, again, as this helps to understand leads and lags in the matching process. Section 3 then addresses the substantial issue of how the demand for statistics might be assessed. This issue is complicated by the global public goods nature of most official statistics. The supply side, and in particular the assessment of statistical capacity, which has received most attention in the literature to date is briefly summarised in Section 4. Section 5 draws together some conclusions. As in the case of any public goods provision there ought to be an explicit regard for willingness-to-pay and the marginal social benefit underlying any new data initiative.

## 2. Some conceptual underpinnings

### *2.1 Some basic concepts*

Statisticians are familiar with basic concepts that distinguish 'data' from 'statistics', although the two terms are frequently used interchangeably. The raw data collected via sample or surveys are usually processed or summarised into statistics. These may be simple summary statistics (means, medians, variances, proportions, Gini coefficients, etc) or they may be more complex measures (such as price indices, national accounts aggregates, etc) forming part of a whole panoply of descriptive measures. It is often useful to distinguish data and statistics per se from 'data constructs', which represent data and statistics in the form of a broader statistical framework or system (Addison *et al*, 1990). The national accounts would be a good example of a data construct, as indeed would be a demographic balance sheet, or indeed a set of poverty profiles. Whilst there is a conceptual distinction between data and statistics (and constructs) the present paper will refer to all forms and formats.

### *2.2 Statistics as commodities*

The conceptual basis for an assessment of the demand and supply of data and statistics rests on the premise of a fundamental understanding about their intrinsic nature as 'goods' (or commodities) in the economic cycle. Statistics are produced by a wide variety of agents and agencies and are used (i.e. demanded) by an equally wide variety of users. The breadth of the variety in both of these

dimensions creates problems in definition, concepts and practical approach in assessing (and matching) the supply and the demand of statistics.

Statistics are produced and used often by different agents (though not necessarily so) and they have all the usual attributes of a 'commodity' found in the economic cycle, in the sense that they are produced and do satisfy wants or needs. The question then arises as to what is the precise nature of this commodity and whether, in turn, this helps us to understand what factors govern its demand and supply. Data and statistics have the hallmarks of an asset class defined in the 2008 SNA called 'intellectual-property products' (ISWGNA, 2008; p 586) which has been subject to subsequent and ongoing revision in the context of intellectual property and research and development (e.g. US BEA, 2013). 'Data and statistics' seem to possess many similar characteristics to music, books or computer software: they are a store of value and they provide a 'stream of benefits to the economic owner … over a period of time' (ISWGNA, 2008; p 586).

At this early stage it is important to recognise the potentially very wide variety of data sources and statistics under our remit. While the statistics referred to in the present context are predominantly official statistics, it is important to recognise that the demarcation between what is 'official' and 'non-official' can change over time and may well differ across countries. Also, in reality the situation is changing rapidly. Non-official statistical sources now play an increasingly important role in information provision, partly due to the explosion in technology (information-gathering is now easy and is applied universally) but also because statistical evidence is now routinely expected and sought in any enquiry where quantification is possible.

### *2.3 Statistics as global public goods*

As far as official statistics are concerned they exhibit all the hallmarks of 'public goods' (Tendulkar, 2009) and, more precisely, of 'global public goods'. Pure public goods possess the properties of non-rivalry and non-excludability. 'Non-rivalry' means that the use by one person does not diminish or preclude the use of that good by another person; and, clearly, official data and statistics that are in the public domain are non-rival. 'Non-excludability' means that it is not possible to exclude anyone from using that good. Clearly, official statistics that are published and freely accessible globally satisfy this property but in those cases where statistics or, more likely, datasets are subject to restricted access (e.g. by a registration procedure or an application process) then they would be more accurately classified as 'impure' public goods, or club goods.

One of the main features of public goods is that outcomes are not economically efficient; which means that they are subject to market failure and, in particular, that users have an incentive to free-ride. One well-known consequence of the free-rider problem is that if goods are produced in a market situation there is likely to be an under-provision. In theory, the efficient level of production would be where the combined marginal rate of substitution (between public and private goods) across all users is equal to the marginal rate of transformation. There are several important consequences of these theoretical results for the production of official statistics. First and foremost is the obvious consequence that one cannot rely on market provision; public provision is necessary. Secondly, as official statistics are *global* public goods (due to their global access and use) and there has to be a global responsibility for their provision. Global users can easily free-ride on nation states' provision (i.e. on the supply of statistics by national statistical offices). Thirdly, whilst the optimal level of provision is reasonably clear in theory it is extremely difficult to ascertain this in practice,

again, this is mainly because of the global nature of the demand for statistics. Hence, statistics may be under-produced (or even over-produced) relative to the optimum position. There is no easy practical check.

In addition to official statistics there is also a wide variety of non-official statistics, ranging from commercial data (businesses, banks etc), research data, and data produced by NGOs and international organisations, etc. In terms of the classification by types of commodity many of these would be classified as either club goods or privately-provided public goods; the reason being that it is easier to prevent free-riders than in the case of official statistics and therefore producers are able to charge for use or access. In the case of marketed non-official statistics it is also potentially easier to assess supply and demand and to set a price at which data are accessed or purchased. But many non-official statistical sources are not marketed; in these cases the data collections (surveys etc) are undertaken to fulfil a particular objective, and the costs of production are met by a budget. If access is granted then other users may be able to free-ride after the primary needs have been met. In the case of non-official statistics the relevant issues may include whether access to these data may be extended and broadened to other users, reducing possible duplication of effort and saving resources in producing similar data via official channels. But there are many other important issues associated with accommodating non-official statistics alongside official statistics. For example, without effective control over data production it is difficult to assess and grade data quality, the coverage might be partial or incomplete, and they tend not to fit well into a comprehensive statistics framework or plan.

It is obviously very difficult indeed to quantify the relative importance of official vs non-official statistics in this area of application. Perhaps one might be able to do this in terms of production cost but this would be quite a tenuous approach. Nevertheless, however we might measure it, official data are probably still substantially more important than non-official data in tracking development processes. It follows therefore that the overwhelming bulk of statistics have the characteristics of global public goods and that standard demand analysis is extremely hard to apply. Furthermore, although standard theory suggests that, because of the free-rider problem, public goods tend to be under-produced relative to the latent demand, in the case of official statistics there is no easy verification of this. Budgets are set by governments and donors, and statistical priorities are usually determined accordingly, usually at the national level but with input from other stakeholders in the international community. So in some instances there could even be an over-production of certain statistics. This might occur when statistics continue to be produced even though they are obsolete or are no longer relevant to needs, perhaps due to administrative inertia, out-dated organisation and practice, etc. This would require investigation and verification: for example, how frequently is the range of datasets reviewed at the national and international level? Finally, there are also cases where countries already have potential data sources but where no statistics are currently being generated[2]

### *2.4 Supply and demand for statistics: leads and lags*

It is of some interest to consider what has been the nature of the leads and lags in the relationship between supply and demand for statistics over time. In general, it is predominantly the case that

[2] For example, civil registration systems, which may be used for validating demographic assumptions on birth and death rates, that are typically assumed from an analysis of census data.

the supply of official statistics is, and has always been, a response to demand, although, as will be seen, there are notable instances where demand has responded to supply. If one traces back to the early European economic statisticians and demographers of the 17th, 18th and 19th centuries (Quetelet, Eden, Graunt, Petty, King, *et al*) the collection of data was in response to enquiries that these empiricists and technocrats initiated. Demographic statistics, rather like national income accounting, originated in France and England in the 17th century, and largely resulted from concerns about the poor and the well-being of the nation. All of this is well-documented (e.g. Bethlehem, 2009; Stone, 1997; Studenski, 1958). The surge in the production of official economic statistics, especially from the mid-20th century onwards, has been driven by a demand for data to chart and monitor economic growth and development, initially in industrialised countries but also in developing countries via their colonial connections. National income, employment and population statistics dominated the supply of official statistics in this early era. Only relatively recently have official statistics expanded into the areas of environment, social statistics (including health), behavioural science and the measurement of living standards. Again this has been largely as a response to national and international initiatives and the need for data in these areas.

In the case of economic and social statistics the growth in the availability of statistics from the mid-20th century prompted the near exponential growth in the study of econometrics and quantitative analysis in the social sciences. As the social sciences became more scientific; statistical techniques were developed to adapt to data supply and to the formats in which official statistics were being made available. (It was usually difficult to acquire additional data - data that were not published or already in the public domain.) So this was possibly an era when demand may have been partly a response to supply. Since about the 1980s, with a marked shift in the development objectives and paradigms away from growth (*per se*) and towards a more direct concern for living standards, poverty and sustainable development, there has been an equally marked shift in the demand for statistics that help us to monitor development outcomes at the micro level and to support the analysis of policy. Micro datasets, longitudinal and multi-purpose household datasets have been generated. These have been produced largely as a result of donor support (even if surveys have been initiated by NSOs) and are a direct response to a demand for these data. Thus, it is now manifestly the case that the demand for data is not predominantly supply-induced. It is the reverse; overwhelmingly, supply has responded to demand. Clearly, with the advent of 'Big Data'[3] and the on-going rapid technological explosion this direction of effect might change yet again and an increasing supply will generate its own demand. But it is the need for particular kinds of statistical evidence in order to monitor development outcomes via MDGs (or similar goals) that is more likely to mean that demand will continue to lead the supply of data.

With demand predominantly outstripping supply, there is a clear need to set priorities and to establish a viable mechanism for determining these priorities. This has usually been achieved via a 'Producer-User' committee, or similar, often under the auspices of the NSOs or an allied ministry. Clearly, there are many stakeholders who ought to be involved: NSOs, national governments, as well as the international community. However, the mechanism for priority-setting, though a crucially important issue, lies outside the immediate scope of this paper.

[3] Large and complex data sets that currently present a major challenge to software and data processing applications.

### *2.5 Supply side trends*

It is self-evident that the production and supply of data has burgeoned in the last four or five decades and at an ever-increasing pace. The most obvious and clear manifestation of this is to note the build-up of statistics and datasets published on the websites of many international organisations over time. The World Bank World Development Indicators are a prime example. On the World Bank Databank[4] statistics are published on over 500 indicators, according to availability, across countries on an annual basis from 1960. Entries for the early years are clearly sparse but the series are extensively populated in later decades. But even a crude indicator such as this conceals many facets of the supply side such as the increasing *range* of statistics (environmental, social and economic) and, especially during the last two decades, the proliferation of micro and longitudinal datasets, together with a focus on increasing openness and a desire for open data.

The two main constraints (or inhibiting factors) on the supply side are, first, statistical capacity (including resource constraints) and, second, the legal restrictions imposed on the dissemination of data. Trends in statistical capacity probably represent the best set of indicators of supply side trends. The measurement, monitoring and appraisal of statistical capacity are aspects that have already received and are continuing to receive significant attention. The efficacy of this measurement will be considered more fully in Section 4.2 below.

The supply side is therefore heavily conditioned by statistical capacity and by existing custom and practice (i.e. the *status quo*). There are already several practical initiatives in place to assess current statistical capacity and to identify how capacity of individual NSOs might be enhanced further. These all relate to the production of official statistics, and are primarily confined to the production that is administered by NSOs even though official statistics may be complied and published by line ministries, central bank, etc. First and foremost has to be the World Bank DECDG database via their ***Bulletin Board on Statistical Capacity*** (BBSC)[5], which is a comprehensive attempt to assemble information at the country level – across all developing countries with summary regional groupings and by lending category. This comprises a summary rating according to statistical methodology, source data and periodicity and timeliness.

A second source of existing information closely related to the enhancement of statistics and of statistical capacity is the set of country ***National Strategies for Development of Statistics*** (NSDS)[6] instigated by PARIS21 (PARIS21, 2013). Judged on the basis of a small sample of country reports, they do provide an excellent platform for a stocktaking exercise, both in terms of the availability, frequency and quality of data produced, and on the human resources and hardware available. The documents are essentially self-assessments, even though they are peer-reviewed. There has been a good take-up of the NSDS by countries (although progress on the peer reviews appears to have been quite slow). Over 80% of IDA countries are currently implementing or designing a strategy, as opposed to the remaining 20% without a strategy or having an expired strategy. The response has been almost as good (75%) for non-IDA (Low and Middle Income) countries. There are of course

[4] http://databank.worldbank.org/data/databases.aspx

[5] http://data.worldbank.org/data-catalog/bulletin-board-on-statistical-capacity

[6] http://paris21.org/NSDS

some incentives for countries' NSOs to comply. By doing so they may attract funds from the World Bank Trust Fund for Statistical Capacity Building (TFSCB).

### *2.6 Demand-side trends*

There are continuing strong demand-side trends in all areas of official statistics. Policy formulation and policy monitoring and analysis are now predominantly evidence-based. This largely drives the inexorable surge in the demand for data by a wide range of users.

Economic data constitute a core demand for development policy analysis. However, while the national accounts, representing the anatomy of the economic system, have been acknowledged as a central organising framework for well over half a century, the system of accounts has continued to evolve and develop. The full internationally-recognised system (SNA[7]) is now quite complex. Many developing countries are as far from being able to implement the full SNA as ever: the targets and the sophistication of the system have advanced at a much faster pace than many NSOs can achieve. There has been much recent criticism of the state of national accounts in African countries (Jerven, 2013) precipitated in part by the upward revision of the estimates of Ghana GDP by over 60%, and the prospect that Nigeria GDP might be revised upwards similarly. Notably, there has been a strong rebuttal and defence of the progress achieved by African statistical offices by Kiregyera (2013) and other members of the African statistical community. But the main point to be made in the present context has to be that, in recent years, the national accounts have been under fire both by users and producers – that is, the demand and the supply sides.

Focussing specifically on the demand side, many development specialists have long questioned the reliance on GDP (and the growth in GDP) and other national accounts aggregates as measures of economic performance (Stiglitz, Sen and Fitoussi, 2009). Of course, this 'new' questioning of a key statistic is not at all new. It has long been recognised that GDP is simply a measure of output of economic activity, and an imperfect one at that. GDP has many well-known deficiencies as a measure of well-being: it is a tenuous link. Long before the Sarkozy Commission report (Stiglitz, *et al* 2009) development economists and analysts increased the range of their demands for data and statistics well beyond national income accounts in order to represent other dimensions of economic and social well-being. This trend began in earnest in the mid 1970s (Chenery, *et al*, 1974) and it prompted a large and extensive multipurpose household survey initiative, the Living Standards measurement Survey (LSMS). This has evolved, via other early initiatives such as the UN Household Survey Capability Programme, into other wide-ranging household surveys, such as the Multiple Indicator Cluster Surveys (MICS), Integrated and Priority Surveys, CWIQ surveys, etc. The LSMS data and survey instruments have been sought after and used in studies of subjective well-being, quality of life and happiness both by economists and psychologists[8]. The bottom line result of this evolving research work is that there is no simple, direct relationship between happiness (however this is measured) and GDP (or gross national income, GNI) and that a range of other factors, non-economic as well as economic, impact on human well-being. This has reinforced the ongoing need for integrated household datasets to supplement the increasing demand for macro and micro level economic data for research and policy analysis.

[7] *System of National Accounts* (ISWGNA, 2013).
[8] For a recent survey of findings see Helliwell, Layard and Sachs (2012).

Two examples of the surge in demand for statistics during the past two decades are as follows. First there has been the need for information to support the MDGs. The demands for data to sustain the indicators have been heavily fuelled by donor support. For example, the MDGs have strengthened the focus on social statistics. Chen *et al* (2013) have suggested that the influence and involvement of donors in survey activities may have helped to side-line the enhancement of some other areas of statistical development, such as the national accounts (see also Lipton, 2013). And the hitherto somewhat unbalanced supply response of NSOs may well have contributed to the call for a rebalancing of effort by Jerven (2011), Devarajan (2013) and others.

A second major demand side trend has been a strong demand for environmental, resource and energy statistics. The relatively recent global concern about global warming and energy supply has brought into sharp focus the many and major lacunae in data supply necessary to meet demand. UNSD (2011) acknowledge that 'monitoring and measurement of progress towards environmental quality is the weakest … pillar of sustainable development' and that 'our capacity to inform about environmental sustainability is severely curtailed by the insufficient production of environmental statistics'. They go on to say that it is not necessarily a lack of data *per se*; it is more a lack of coherent and coordinated national programmes with regular delivery of relevant statistics and indicators.

## 3. Demand for statistics

### *3.1 Identifying the users*

What are the specific motivations that underlie the demand for official statistics and from where do these demands emanate?

Focussing specifically on development objectives there are three clear overall motivations:

(a) to provide the statistical evidence for monitoring progress towards development goals;

(b) to generate data and statistics necessary to underpin evidence-based policy-making;

(c) to generate data and statistics to underpin research – i.e. this is not just for observation and tracking, it is the need for empirical evidence to further our understanding of socio-economic mechanisms and the process of development.

These are not separable or independent motivations; but (a) and (b) have clearly dominated (c) in prompting the production of data and statistics by NSOs. If we identify who are the drivers in the demand for official statistics then the reasons are clear.

The main agents/institutions that create these demands are:

(i) national and regional governments (including institutions of government – line ministries, central bank, etc);

(ii) international organisations (including their affiliates, development agencies, etc) who are also principal donors for much of the technical assistance, including

survey work;

(iii) NGOs;

(iv) academic/research institutions (individuals, universities, and other research centres);

(v) corporate enterprises (including multinationals), consultancies.

But the statutory responsibilities for producing official statistics, and funding, rest with governments (i) and international organisations (ii). The other three groups of institutions are not principal drivers in the production of statistics, and certainly not in official statistics. This is primarily because they do not generally provide funds nor do NSOs have any statutory responsibilities to supply statistics to them. Funding is a necessary (though obviously not a sufficient) condition for the production of official statistics, so it is inevitable that research needs (c) may be side-lined in the process. Note, however, that there are major issues with regard to funding and to the source of funds in particular. Funds should not be from sources that reflect the vested interests of donors otherwise they undermine the integrity of the data and statistics generated.

The recognition that there are different groups of users is a touchstone to the important issue of establishing effective communication channels between users and producers, and responding to the needs of users. This has long been recognised as being necessary and important but it is very difficult to implement. Kiregyera (2013) refers to an increasing awareness of the 'statistical value chain' by producers in Africa, and what he perceives to be more emphasis on data analysis and a better collaboration with users and data analysts. Again in the Africa context he refers in particular to an increase in the 'release of micro data to serious researchers' (Kiregyera, 2013: p 16). This has been supported and coordinated via the Accelerated Data Program (ADP) that is hosted by PARIS21, for countries that have joined. Of course this begs the question as to how 'serious' researchers are identified. The different categories of user have varying rights of access to data: line ministries would have better access than academic institutions and international organisations probably have better access than anyone, especially if they provide the necessary funding.

### *3.2 Assessing the demand*

Clearly, any attempt to establish the extent of the demand for statistics it is likely to be an open-ended exercise for two main reasons. First, as we have seen, users of official statistics are either not usually charged a fee for their use of data or, at least, the charge is not commensurate with the (marginal) cost of production. Because users do not face a budget constraint demand is unconstrained it is therefore inevitable that the demand for data will exceed any potential supply. Secondly, it is simply not possible to anticipate data needs for research beyond a relatively short time horizon. A good example of this is that it would not have been easy to anticipate the scale of the current use of large multipurpose household surveys 25 years' ago. Nor would it have been possible to envisage the growth in the supply and use of environmental statistics prior to the setting up of the International Panel on Climate Change (IPCC) in 1987 and the Kyoto Protocol in 1997. There are several other examples of where major unanticipated changes in the demand for statistics have occurred in recent decades but all show changes that are *increases* in demand, very rarely have demands *decreased*. What is clear is that the trend towards the demand for more data, better data,

more frequent data, more timely data, etc, is ever upwards. As data and official statistics are provided to users at no cost or low cost demand is likely to be insatiable: demand is always likely to outstrip supply. Nevertheless it is important for NSOs to be continually aware of the data needs of users, not only by government users and international organisations, but also for NSOs to be integral partner in research efforts more generally.

A central challenge for all data producers, national or international, is to assess and take stock of the likely demand for different kinds of data and statistics. The supply of official statistics are no longer exclusively geared to the demands by government to serve country needs, or influenced by international organisations, so the potential needs of all categories of user have to be taken into account. The basic problem to be tackled is to assess the latent demand for data in a rapidly changing environment. Latent demand and its closely-related counterpart, supply-induced demand, are concepts that have been used in travel and transportation, and in health economics. These concepts may prove helpful in assessing the potential demand for new data of different kinds. But first, what are these concepts and how might they relate to the demand for data?

*Latent* demand is simply defined as the pent-up demand for a good or service that is desired but unrealised because of constraints (Mokhtarian, 2004). This is a clear instance of unfulfilled demand. For example, in the case of transportation studies there is often a desire to assess the potential traffic demand on, say, a new highway or facility, perhaps as part of the project appraisal. A closely related concept is *induced* demand. Induced demand is defined as the realised demand that is drawn out (i.e. 'induced') by additional capacity (or by the relaxation of constraints). Standard examples of induced demand occur in the case of transport (e.g. a new highway) or health (e.g. new treatments).[9] Both of these concepts seem to relate directly to our problem of estimating the demand for data. In the first place there is clearly a latent demand for data, in the sense that users would like data that are not yet produced; and, in the second place, if new data are produced then this might induce a further demand for it.

As noted above, there exists a large literature on attempts to estimate latent and induced demand in the applied areas for transport and health (Lee, Klein and Camus, 1999). It would seem that the main approach has been to estimate demand equations based on exogenous explanatory variables. In the case of traffic demand (in terms of traffic volume) the exogenous variables include, land use, population, income and employment. Estimated equations are then used to 'predict' traffic volumes in other situations. Clearly this approach is not applicable to the problem of assessing the latent demand for data. We do not face a set of 'observations' based on, say, individuals or households. The units creating the demand for data are quite heterogeneous: they may be as small as individuals or as large as institutions. And the best one can hope to achieve is to draw a sample of potential users that is a representative cross section and to explore the latent demand based on this sample. It is therefore not yet clear how formal demand analysis can help.

(a) *Unclassified, unranked data demand inventory*

[9] In the area of health economics there is considerable literature describing 'supplier-induced demand', which would appear to be a quite different concept. For example, the 'patient-doctor (i.e. physician)' relationship is a classic 'principal-agent' problem in the sense that there is asymmetric information between principals (i.e. patients) and agents (i.e. physicians) and where there may be circumstances where physicians might prolong the treatment time beyond the patient's needs (Ngyuyen, 2011).

The simplest and most basic approach would be to devise a survey (by interview or by questionnaire) in which a cross-section of potential users are asked to report/suggest data needs by topic - and to include the frequency, format, and the use (actual current or intended) to which the data would be put. Thus, the survey would cover data needs not currently being met as well as anticipated future needs. In order to encourage users to be mindful of resource and production costs it should cover questions about the availability of possible funding. Of course, stakeholders may not be considered 'equally important, and this requires a judgemental decision. Obviously, the problem with asking users to submit unbounded responses is that there is no easy way of testing the strength of demand.

(b) *Classified and ordinal response inventory*

A variant of the above approach would be to invite respondents to grade their responses on a 'Likert-type' scale, such as 'strongly required', 'desirable', 'of some interest', etc. This would yield information about users' perceived data priorities.

It may well be that an informed market-research approach to an assessment of the latent demand for statistics could prove fruitful. A major difficulty with this or any other formal approach is to identify a suitable sampling frame. Many users and potential users are diffuse and are not readily identifiable. Devising a suitable questionnaire is likely to be much less problematic. It should be relatively easy to survey government (i): including line ministries, the central bank, etc, and international organisations (ii), and the major NGOs (iii). However, as regards the multitude of academic and research organisations, individuals and private sector users (iv) and (v), the most effective initiative might be to hold forums at conferences and to invite submissions from professional societies and organisations. Research grant funding bodies should also be approached; they should have good insights into changing data needs.

## 4. Supply of statistics

### *4.1 Identifying the producers*

By definition, official statistics are produced by government agencies (NSOs, line ministries, regional and local government agencies) or other public bodies including international organisations. The activities of NSOs range well beyond data production and assembly of statistics. Their activities also include dissemination (publication and statistical information services), coordination (between producers, to avoid duplication of effort), technical advice and consultation, training (including computer training), and communication with users. Thus, unlike users, the producers should be relatively easy to identify. So the principal issue is to determine the ability or capacity of countries to produce, deliver and service official statistics – so as to embrace all of these activities.

### *4.2 Measuring statistical capacity*[10]

Statistical capacity has been defined in various ways—there is no universally accepted definition in the literature—and this has caused some basic confusion and continuing debate. However, it is central to any appraisal of supply side issues to have some way of assessing statistical capacity both in a country and, ultimately, globally. Put very simply, at the two extremes, the capacity to produce

[10] This section is derived from material previously included in Round (2012)

statistical information can be viewed either from an input or an output perspective. Traditionally, the input perspective has been equated with 'statistical capability'—that is, it is an attempt to assess the resource inputs needed to produce statistics. So in practical terms this means measuring and monitoring changes in human and financial resources, equipment and facilities, etc. An alternative is to consider the output perspective, emphasizing effective or 'realised statistical capacity'—that is to consider the quantity and quality of statistical output. Usually this is assessed by the availability of specific statistical data—e.g. time series that meet given quality standards, etc.

Ngaruko (2008a) argues however, that, within the context of the PARIS21 results-chain framework (Laliberte, 2002), statistical capacity ought to be viewed as determining the *resources* available to carry out statistical activities—that is, it should be considered explicitly from the input side. His suggested list of resources includes:

- human resources (technical, administrative, support, plus data-producing agency staff);
- infrastructure (buildings, power, etc.);
- human resources management practices (hiring, firing, promotion, training, etc.);
- finance and its characteristics (level, sources, stability);
- computing facilities (availability, maintenance, and updating of IT infrastructure);
- transport, communication and office equipment;
- statistical practices and the regulatory framework.

Indeed, a full application of the PARIS21 results-chain would generate assessments of the level and change in all these resources and, of course, much more.

An alternative to this is the 'output' approach represented by the World Bank DECDG statistical capacity indicator(s), although, as we note subsequently, there are some other important technical differences between this and the PARIS21 results-chain approach. DECDG (World Bank, 2008) has created a measure of statistical capacity via a set of selected indicators. DECDG has been rating countries worldwide in terms of their statistical capacity since 2004. The approach has been to develop a composite indicator of statistical capacity based on publicly-available information in three particular dimensions:

(i) statistical methodology

(ii) source data

(iii) periodicity and timeliness.

Quoting DECDG (World Bank 2008), the first dimension, *statistical methodology*, measures a country's ability to adhere to internationally recommended standards and methods. The second dimension, *source data*, reflects whether a country conducts data collection activities in line with internationally recommended periodicity, and whether data from administrative systems are available and reliable for statistical estimation purposes. The third dimension, *periodicity and timeliness*, captures the availability and periodicity of key socioeconomic indicators (nine of which are MDG indicators).

In the DECDG statistical capacity indicator s*tatistical methodology* is measured with 10 indicators; *source data* by 5 indicators; and *periodicity and timeliness* by a further 10 indicators. Within each dimension a set of indicators are constructed on the same scale (0 to 1) and are then combined with equal weights. A composite statistical capacity indicator is then calculated as a simple arithmetic mean of the three dimension indicators. There have been only minor revisions to the indicators since their inception; around 2008 the three dimensions were renamed (as above) and there has been some refining of the scores on one sub indicator ('access to water'). A key advantage is that the indicators do not require data collection from countries; they are based on existing databases of international organizations. The indicators are calculated and published annually by country and by region.

Ngaruko (2008a) makes two criticisms of the DECDG statistical capacity indicators. First, as already mentioned, the 'dynamics' of statistical resources are multifaceted, so that changes in statistical capacity (which in turn affects the level at which statistical activities can take place) can arise from any or all of the resource inputs, and these are not captured by the index. Ngaruko is therefore emphasizing that it is essentially an output and not an input measure.[11] Second—and this is Ngaruko's central point—the DECDG indicator overlooks *capacity utilization*. So a country can increase its statistical activity and output by using dormant capacity rather than acquiring fresh capacity. The aid effectiveness implications are clear: if the purpose of aid is to enhance statistical capacity then it might not show up in the form of increased activity or outputs until the enhanced capacity is utilized. Ngaruko goes further by suggesting that this might be a major contributing factor to the volatility of the index—although this argument is by no means clear. It could be argued, for example, that the existence of under-utilized capacity might smooth out changes in statistical activity and data production.

Whilst overall statistical capacity indicators are one way of assessing the ability of country NSOs to carry out their statistical activities and services, there are alternatives. A revealing alternative is reported by Chen, Fonteneau, Jütting and Klasen (2013; Figure 1). This chart, reproduced from the MDG database, shows the availability of MDG data for selected indicators (percentage of 174 developing countries) recorded in four intervals between 2000 and 2012. Chen et al (2013) report, positively, that the MDG data gap is reducing – but there is (a) a huge variation in data gaps across indicators and (b) the data gap is still very wide across the board. So there is unquestionably some way to go in fulfilling needs from the supply side.

A major constraint in enhancing statistical capacity is funding. Funds are either channelled through the World Bank's Trust Fund for Statistical Capacity Building (TFSCB), regional development banks (such as the African Development Bank), European Commission, or from bilateral donors (U.K., Denmark, etc). Significant support has come indirectly via other major projects, such as the International Comparison Program for Africa (ICP-Africa). Much of the work on the supply side has therefore focussed on evaluations of this aid effort, especially in terms of the allocative efficiency of this funding both across countries and by purpose (OPM, 2009).

[11]. Ngaruko (2008b) indicates that 19 of the 25 component indicators relate to statistical activities and outputs.

## 5. Conclusions

In a wide-ranging article Sanga (2013) has noted the data gaps and data deficiencies in the realm of economic and social statistics in Africa. While it is possibly most acute of all in Africa the problem is more general than this. It is a global problem and it is also a problem that is manifest in fields other than economic and social statistics, such as in environmental and demographic statistics. There is an assumption that the demand for statistics always exceeds supply, but we have noted that, historically, this has not always been the case and it might not even be true in all fields today. It is possible that, through inertia and some protection of the *status quo*, there are data surplus areas and therefore that efforts could be redirected towards deficiencies in other areas. These possibilities have not been pursued here though they could be examined further.

We have argued that data and statistics are a produced asset and therefore that they are subject to the usual forces of supply and demand in much the same way as any other commodity. However one of the features that cause us most difficulty is that official statistics are a global public good. On the demand side it means that as there is usually no charge for accessing data, users (or potential users) have no perception of its cost, or even of the opportunity cost of producing it. Also, it means that demand for many kinds of statistics is global and this creates particular difficulties in identifying the level and extent of demand. The problem is one of assessing latent demand. We have examined some of the literature relating to estimating latent demand for other kinds of public goods. There could be parallels with how we assess the demand for, say, a new highway. A highway is a physical asset whereas statistics are more akin to intellectual property. Nevertheless it might be possible to adapt existing methodologies: this is yet to be determined. The supply side, and in particular statistical capacity building, has received most attention to date. While this is also non-trivial and is beset by many conceptual difficulties it seems to be more tractable than the demand side.

One important area that has not been tackled in this paper is the matching of demand and supply for data and statistics. Historically there have been periods of supply-led statistical development as well as the more usual demand-led periods. In a current context Scott (2005) has characterised some countries as being demand-constrained, whilst making the case for a strengthening of the evidence base for policy-making in developing countries. This is an area that certainly needs considerably more work.

## References


Addison, T, L Demery, M Ferroni, J I Round (1990) *Making Adjustment Work for the Poor: A Framework for Policy Reform in Africa*, The World Bank, Washington D C.

Bethlehem, J (2009) 'The Rise of Survey Sampling', Discussion Paper 09015, Statistics Netherlands, The Hague

http://www.cbs.nl/NR/rdonlyres/BD480FBC-24CF-42FA-9A0D-BBECD4F53090/0/200915x10pub.pdf

Chen, S, F Fonteneau, J Jütting and S Klasen (2013) 'Towards a Post 2015 Framework that Counts: Aligning Global Monitoring Demand with National Statistical Capacity Development', PARIS21 Discussion Paper Series No 1, OECD Paris.

http://mortenjerven.com/wp-content/uploads/2013/04/Panel-8-Jutting.pdf

Chenery, H, M S Ahluwalia, C L G Bell, J H Duloy and R Jolly (1974) *Redistribution with Growth*, Oxford University Press for The World Bank and Institute of Development Studies.

Devarajan, S (2011) 'Africa's Statistical Tragedy', Keynote speech at the IARIW-SSA conference on "Measuring National Income, Wealth, Poverty and Inequality in African Countries.

http://blogs.worldbank.org/africacan/africa-s-statistical-tragedy

ISWGNA (2013) *System of National Accounts 2008*, Interagency publication, New York, Paris.

Helliwell, J, R Layard and J Sachs (editors) (2012) *World Happiness Report*, The Earth Institute, Columbia University, CIFAR and Centre for Economic Performance, London

Jerven, M (2013) *Poor Numbers: how we are misled by African development statistics and what to do about it,* Cornell University Press.

Kiregyera, B (2013) 'The Dawning of a Statistica Renaissance in Africa', presented at the conference African Economic Development: Measuring Success and Failure, Vancouver, November 2013.

http://mortenjerven.com/wp-content/uploads/2013/04/AED_Panel_8-Kiregyera.pdf

Laliberte, L (2002) *Statistical Capacity Building Indicators*, *Final Report*, PARIS21, OECD, Paris.

http://paris21.org/sites/default/files/scbi-final-en.pdf

Lee, D, L Klein and G Camus (1999) 'Induced Traffic and Induced Demand', *Transportation Research Record*, 1659, 68-75.

http://www.worldbank.org/transport/roads/rpl_docs/apbinduc.pdf

Lipton, M (2013) 'Africa's National Accounts Mess: A Review Article', *Journal of Development Studies*, DOI: 10.1080/09662839.2013822668.

http://www.tandfonline.com/loi/fjds20

Mokhtarian, P L (2004) 'Understanding the Latent Demand in Traffic', PPT presentation at Caltrans Division of Research and Innovation.

http://www.dot.ca.gov/researchconn/past_speakers/DrMokhtarian/index.htm

Ngaruko, F (2008a) 'The Word Bank's Framework for Statistical Capacity Measurement: Strengths, Weaknesses and Options for Improvement', *African Statistical Journal*, 7:149-69.

http://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/8-worldbankframe.pdf

Ngaruko, F (2008b) 'The World Bank's Approach to Statistical Capacity Development: The Missing Link', African Statistical Newsletter, 2(2): 16-18.

http://ecastats.uneca.org/acsweb/Portals/0/African%20Statistical%20Newsletter/ACSNewsletterJune2008_2.pdf

Ngyuyen, Ha (2011) 'The principal-agent problems in health care: evidence from prescribing patterns of private providers in Vietnam', *Health Policy and Planning*, 26(1): 153-162.

http://heapol.oxfordjournals.org/content/26/suppl_1/i53.full

OPM (2009) Evaluation of the Implementation of the Paris Declaration: Thematic Study – Support to Statistical Capacity Building, OECD Paris and DFID London.

http://www.oecd.org/dac/evaluation/dcdndep/43189996.pdf

PARIS21 (2013) *National Strategies for the Development of Statistics Progress Report* March 2013, OECD, Paris.

http://paris21.org/sites/default/files/NSDS_Status_Table_Mar2013.pdf

Plunkett, S (2011) 'Untangling Intangible Assets', OECD Observer, No 285.

http://www.oecdobserver.org/news/fullstory.php/aid/3556/Untangling_intangible_assets.html

Round, J I (2012) 'Aid and Investment in Statistics for Africa', ReCom Working paper WP093, 2012/93, UNU-WIDER, Helsinki.

http://www.wider.unu.edu/publications/working-papers/2012/en_GB/wp2012-093/

Sanga, D (2013) 'The Challenges of the Narrative of African Countries' Development: Data Demand and Supply Mismatches', presented at the conference African Economic Development: Measuring Success and Failure, Vancouver, November 2013.

http://mortenjerven.com/wp-content/uploads/2013/04/AED_Panel_8-Sanga.pdf

Scott, C (2009) 'Measuring Up to the Measurement Problem: the Role of Statistics in Evidence-Based Policy Making', PARIS21 Secretariat, OECD Development Cooperation Directorate, Paris.

http://paris21.org/sites/default/files/1509.pdf

Stiglitz, J, A Sen and J-P Fitoussi (2009) *Report by the Commission on the Measurement of Economic Performance and Social Progress*, Insee, OFCE, and OECD, Paris.

http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf

Stiglitz, J (2009) 'The Great GDP Swindle', *The Guardian*, 13 September 2009.

http://www.theguardian.com/commentisfree/2009/sep/13/economics-economic-growth-and-recession-global-economy

Stone, J R N (1997) *Some British Empiricists in the Social Sciences 1650-1900*, Cambridge University Press.

Studenski, P (1958) *The Income of Nations: Theory, Measurement, and Analysis: Past and Present*, New York University Press.

Tendulkar, S (2009) 'Demand for Better Statistics and Use of Data', PARIS21 Secretariat, OECD Development Cooperation Directorate, October.

http://paris21.org/sites/default/files/datause-Tendulkar-Consortium09.pdf

UNSD (2011) *Environmental Statistics and Sustainable Development at Rio+20*, United Nations Statistical Division/DESA.

http://unstats.un.org/unsd/environment/Rio+20.pdf

U.S. Bureau of Economic Analysis (2013) 'Changes to How the U.S. Economy is Measured Roll out July 31', BEA News July 23 2013.

http://blog.bea.gov/2013/07/23/gdp_changes/

World Bank (2008). 'A Note on the Statistical Capacity Indicator', DECDG, World Bank, Washington.

http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/EXTWBDEBTSTA/0,,contentMDK:22260537~menuPK:6345105~pagePK:64168445~piPK:64168309~theSitePK:3561370,00.html#Statistical_Capacity_Indicator