MEASUREMENT

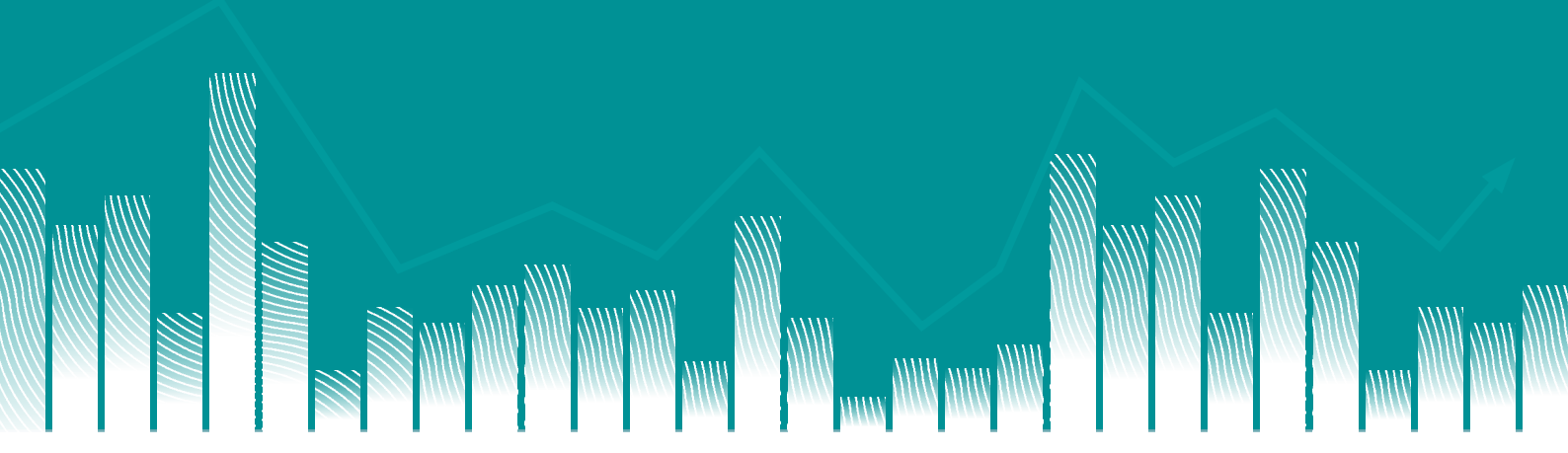
MEASURING STATISTICAL CAPACITY DEVELOPMENT:

A REVIEW OF CURRENT PRACTICES AND IDEAS FOR THE FUTURE

FRANÇOIS FONTENEAU, BARBARA YAEL BAREDES AND CHARLOTTE MAYARD

PARIS21 DISCUSSION PAPER N°. 11

MARCH 2018



PARIS21

Partnership in statistics for development in the 21st century

PARIS21 Discussion Paper, No. 11 March 2018

Measuring Statistical Capacity Development:

a review of current practices and ideas for the future – moving towards Statistical Capacity 4.0

This paper was prepared by François Fonteneau, Barbara Yael Baredes and Charlotte Mayard of PARIS21. Additional inputs were provided by Richard Roberts and José-Antonio Mejia.

PARIS21
Partnership in statistics for development in the 21st century

PARIS21 Discussion Papers

The PARIS21 Discussion Paper series features expert authors ranging from statisticians to academics who seek to lend their voices to the discourse on statistical capacity development.

For more: http://paris21.org/paris21-discussion-and-strategy-papers

Please cite this publication as:

PARIS21 (2018), "Measuring Statistical Capacity Development", PARIS21 Discussion Paper, No. 11, Paris. http://paris21.org/paris21-discussion-and-strategy-papers

CONTENTS

EXECUTIVE SUMMARY 4

1. INTRODUCTION 5

2. EXPLORING EXISTING ASSESSMENTS 6

a) Rationale 6

b) The Capacity Development 4.0 Framework 6

c) Selection of assessments 7

d) Classification of assessments 8

e) Methodology 10

3. HOW CURRENT ASSESSMENTS MEASURE STATISTICAL CAPACITY 12

a) Overall results 12

b) Results by classification 13

c) Individual results 15

d) Overlooked or under-tackled dimensions 17

e) Statistical capacity in current assessments 18

4. DIVING DEEPER INTO SELECTED DIMENSIONS 19

a) Organisational skills and knowledge: “Statistical production processes” 19

b) Questions and indicators for “orphan” categories 20

5. CONCLUSION 29

ANNEX 1: LIST OF QUESTIONNAIRES 30

REFERENCES 32

Executive summary

The paper provides evidence on what 14 of the most referenced assessments of statistical capacity aim to measure, and how they do so by comparing them with the dimensions of the PARIS21 Capacity Development 4.0 (CD4.0) framework. This paper highlights the dimensions that are mostly covered and those that are neglected by the reviewed assessments. It discusses and explains when and why this could bias our understanding of statistical capacity, therefore, the associated actions for developing capacity. It provides an in-depth analysis of one of the most tackled dimensions (statistical business processes), and suggests new indicators to address those that are currently not measured, but which were deemed relevant by the group of experts on the Task Team.

Summarising the results, as a group the assessment tools characterise statistical capacity as predominantly composed of skills and knowledge at the organisational level (specifically, statistical production processes, quality assurance and codes of conduct) and of resources at the systemic level (legislation, principles and institutional setting as well as existing data). There is very little coverage of incentives at any level; of skills and knowledge at the systemic level (data literacy and knowledge transfer); or of the capacity at the individual level in general. This reveals that the understanding of statistical capacity is only partial. While the assessment tools vary according to type, method of administration, and purpose, they nevertheless portray a similar definition of statistical capacity.

Future lines of work may include further analysis of the assessments' questions for similarities, differences and duplications[1]. Ultimately, the analytical question bank created for the purposes of the Task Team (Open Assessment Repository, available online at: oar.paris21.org) could serve as a reference for designing harmonised and comprehensive assessments, with the double objective of generating meaningful quality data while reducing the burden on respondents.

This paper contributes to the PARIS21 "Capacity Development 4.0" (CD4.0) Task Team. It directly contributes to the second objective of the Task Team, to look into ways CD4.0 needs to be measured. The Authors are grateful to all Task Team Members whose contributions are dully recognised.

[1] Although indicative, the distribution of questions from the assessment tool(s) across the conceptual framework categories does not allow for a complete comprehensive view – this would require a semantic analysis. A small number of questions may be appropriate for less complex categories.

1. Introduction

The 2030 Sustainable Development Agenda, with its 17 Sustainable Development Goals (SDGs) and framework of 244 indicators for monitoring and evaluation, is creating an unprecedented demand for data. This poses new challenges and opportunities for national statistical systems (NSS). These demands are compounded by the local data requirements from national and sectoral policy frameworks and the needs of civil society. Despite recent improvements, NSSs in many countries, in particular the poorest, are struggling to produce basic data for policymaking and programme implementation. The emerging data demands of the SDGs add to this challenge in a financially constrained environment.

In this context, capacity development in data and statistics has become a widely recognised priority, as stated in the Cape Town Global Action Plan for Sustainable Development Data (CTGAP), unveiled at the UN World Data Forum in January 2017. Following its 2017 Annual Meetings, PARIS21 formed a task team to revisit statistical capacity development with the purpose of aiding with the implementation of the CTGAP. The "Capacity Development 4.0" (CD4.0) Task Team had three objectives: first, to propose a conceptual framework for CD4.0; second, to look into ways CD4.0 needs to be measured; and third, to design a number of implementation principles.

This paper contributes to the task team's measurement objective by presenting the results of a review undertaken by the PARIS21 Secretariat. These results were discussed during the Workshop on New Approaches to Statistical Capacity Development, organised jointly by PARIS21 and the United Nations Development Programme (UNDP) in Paris on 11-12 December 2017.

The review focuses on the quantifiable aspects of measuring capacity, and explores (i) what is currently being measured, and (ii) how is it being done. Fourteen assessment tools were analysed to evaluate their relevance for measuring the various dimensions of statistical capacity. The CD4.0 conceptual framework was used as a reference for this review, which in turn helped refine the framework. The framework lists all of the dimensions and categories, which need to be covered to when measuring capacity in a way that is compatible with CD4.0. A public online platform, the Open Assessment Repository (OAR),[2] has been developed to compile all the assessments and map their content to the framework. All the data used to support this research can be found on the OAR.

[2] www.oar.paris21.org.

2. Exploring existing assessments

a) Rationale

The starting point of our analysis was the absence of a systematic approach to defining capacity as a basis for designing assessments and capacity development programmes. There have been criticisms in the past of the way in which statistical capacity is conceptualised in international assessments – there is a tendency to emphasise activities and outputs, while neglecting the utilisation of capacity (e.g. a country may have the ability to conduct a survey but not do it) (Ngaruko, 2008).

There are two main consequences of this lack of a systematic approach. One of them is the proliferation of assessments by international organisations. This places a significant response burden on countries – paradoxically constraining those with low capacity even further. The other is the repetition of topics, areas, and even indicators/questions across many of the assessments. The majority of them focus solely on the most tangible aspects of capacity (mainly methodology, resources and statistical laws).

b) The Capacity Development 4.0 Framework

Following the CD4.0 framework, statistical capacity can be defined as "the ability of a country's national statistical system, its organisations and individuals to collect, produce, analyse and disseminate high quality and reliable statistics and data to meet users' needs" (PARIS21, 2018: 4).

The framework identifies three levels (individuals, organisations and system) and five targets (resources, skills and knowledge, management, politics and power and incentives). The intersection of a level and a target is called a category (e.g. individual resources), and the components of a category are called dimensions (e.g. professional background). Table 1, extracted from *Proposing a Framework for Capacity Development 4.0 (draft)* (PARIS21, 2018) provides further detail. For more information about the framework, please consult the paper.

Table 1**: Capacity Development 4.0 framework**

Target/Level	Individual	Organisational	System
Resources	Professional background	Human resources Budget Infrastructure	Legislation, principles and institutional setting Funds infrastructure Plans (NSDS, sectoral…) Existing data
Skills and knowledge	Technical skills Work 'know-how' Problem solving and creative thinking	Statistical production processes Quality assurance and codes of conduct Innovation Communication	Data literacy Knowledge sharing
Management	Time management and prioritisation Leadership	Strategic planning and monitoring and evaluation Organisational design HR management Change management Fundraising strategies	NSS co-ordination mechanisms Data ecosystem co-ordination Advocacy strategy
Politics and power	Teamwork and collaboration Communication and negotiation skills Strategic networking	Transparency Workplace politics	Relationship between producers Relationship with users Relationship with political authorities Relationship with data providers Accountability
Incentives	Career expectations Income and social status Work ethic and self-motivation	Compensation and benefits Organisational culture Reputation	Stakeholders' interests Political support Legitimacy

c) Selection of assessments

We identified and reviewed 14 assessment tools that are currently being used to measure and gauge various aspects of national data and statistics, including the capacity and performance of statistical institutions and the quality of their outputs - sometimes loosely referred to as "statistical capacity" by their authors or users. The 14 assessments are the following (see Annex 1 for further details of these tools):

1) Assessing the National Health Information System. An Assessment Tool from HMN and WHO (HIS)
2) African Statistical Development Indicators from UNECA (StatDI),
3) In-Depth Country Assessment from FAO (IdCA),
4) Data Quality Assessment Framework for National Accounts Statistics from IMF (DQAF for National Accounts),

5) Environment Statistics Self-Assessment Tool from UNSD (ESSAT),
6) Generic National Quality Assessment Framework from UNSD (Generic NQAF),
7) Global Assessment of the National Statistical System from Eurostat, UNECE and EFTA (GA),
8) Light Self - Assessment Questionnaire on the implementation of the European Statistics Code of Practice from the European Commission – EC/Eurostat (Light SAQ),
9) Pan-African Statistics Programme: Peer Reviews of NSIS/NSSS in African countries from Eurostat and AUSTAT (PAS),
10) Self-Assessment Guidance Questionnaire from UNECA (SAGQ),
11) Snapshot (Eurostat),
12) Statistical Capacity Indicators from the World Bank (SCI),
13) Tool for Assessing Statistical Capacity from US Census Bureau (TASC)
14) the extra modules added by IDB to the Tool for Assessing Statistical Capacity (TASC v.IDB).

Although not all of them are specifically meant to measure capacity, their inclusion in this analysis proved relevant because they are often referred to as "capacity assessments questionnaires", and provide key input to the design or monitoring and evaluation of statistical capacity development programmes. Their scope and structure is similar to the capacity assessment tools.

The unique purpose(s) of each tool and their contribution to the measurement of CD4.0 were recognised and documented. The selection includes, in addition, tools developed for voluntary use by countries and those used by international organisations with the aim of comparing countries' statistical development or compliance with international standards.

We used the CD4.0 framework as the basis for categorising the questions/indicators contained in these assessment tools. We then proceeded to identify which categories, levels and targets of capacity development are the most tackled by these assessments, and which are the least represented. This exercise allowed us to identify overlaps and similarities between questions/indicators from different assessments.

d) Classification of assessments

In order to control for possible variations across subtypes of assessments, we decided to classify them and test for differences. We used three criteria to group them: by type of instrument, by administration method and by purpose. These classifications are non-exclusive; therefore, some assessments may fall into more than one of the groups.

Assessments can be initially classified by **type of instrument**. There are three groups: the most frequent one is *structured questionnaires*, such as the TASC from the US Census Bureau. A second type is *open guidelines*, such as the Light SAQ. A third involves data collection by means of *secondary information sources*, such as the SCI; this type focuses on inferring capacity based on reported methodologies and outputs. Among the 14 tools, there are 10 structured questionnaires, 3 open guidelines and one secondary data collection exercise.

Assessments can also be grouped by **administration method**, of which we have identified three types:

1) *Self assessments* (administered by the authorities of the country under assessment). An example is the Generic NQAF.
2) *External assessments* (administered and facilitated by external experts from international or regional organisations). Some of these assessments can involve actual data collection on the field (most cases, in countries); while others are typically done without such as the Statistical Capacity Indicator from the World Bank.
3) *Peer reviews* (conducted by an external evaluator from another country with experience in the same field). An example is the PAS.

These three groups can be further mapped to the typology of NSS assessment tools proposed by PARIS21 in its "Guide to Assessment of National Statistical Systems" (Strode, 2017) that distinguishes four types according to their **purpose**.

The tools of type A, "National Planning and Advocacy for Statistical Improvements and Modernisation", are meant to be used by national authorities for diagnosing the NSS with the aim of developing an action plan for improving statistical performance. They are either self-completion tools or are used by international experts or regional peers.

Type B tools, "Development Partner Project Design and Monitoring", are used to advise development agencies and their national partners on the strengths and weaknesses of statistical systems, and as a means of designing programmes or projects to address some of the issues identified. They are usually completed and analysed by experts, or completed by multiple national stakeholders and then analysed by experts.

The tools of type C, "Global Monitoring of statistical performance", are intended to monitor statistical performance at an international level, and are applied by international agencies without direct country participation. They are mostly based on publicly available data from national statistical websites or from survey returns to international bodies.

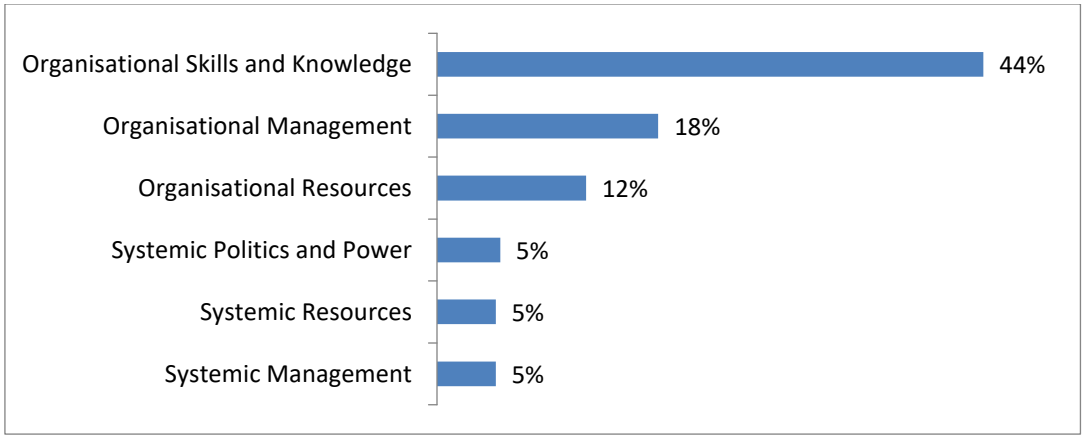
Type D tools, "Data quality assurance and compliance with codes of practice, norms and standards", are usually mandatory and are used by international or regional bodies to assess compliance with their codes of practice and quality standards. They can also be non-mandatory and used for advisory purposes. They are mostly applied by experts or peers (Strode, 2017). The detailed classification is in Annex 1.

e) Methodology

We developed the Open Assessment Repository (OAR)[3] and populated it with all the questions from the 14 assessment tools listed earlier, linking them to the categories and dimensions of the conceptual framework (Table 1). Each question was associated to one (or in some cases two) dimension[4]. The aim was to observe which categories and dimensions of the conceptual framework are tackled by each assessment. The repository contains 1971 questions, but taking account of the cases where a question corresponds to more than one dimension, a total of 2407 associations exist. All calculations in this paper were based on this number.

As an example, the TASC from the US Census Bureau comprises a total of 229 questions, 44% of which belong to the organisational skills and knowledge category and 18% to the organisational management category (Figure 1). Remarkably, it contains no questions under the categories of individual resources, individual management, individual politics and power, or systemic skills and knowledge.

Figure 1: **Relative frequency distribution of categories in the TASC**



[3] We also developed a metadata model for the assessment tools that will be available on the OAR. It is designed to give users ready access to relevant information about each tool, to help them decide if it suits their purposes. The model provides for recording more than 30 elements of information about each assessment tool, such as identification details.

[4] Three people worked independently on each assessment, and held meetings to align their coding criteria. After the initial coding and data-entry, a final quality control was performed for each dimension. The coding was done once before the dimension was commented by the Task Team, and a second time including the modifications suggested by members. The results in this paper are those of the second round.

Overall, 80% of TASC focuses on the organisational level, with 48% of the questions assessing skills and knowledge.

Noticeably, five dimensions represent almost 60% of the TASC questions, all belonging to the organisational level. These dimensions are: statistical production processes (28%), strategic planning and monitoring and evaluation (9%), quality assurance and codes of conduct (8%), communication (8%) and HR management (6%).

The following 13 dimensions are not covered by the TASC:

- **at the individual level**: Time management and prioritisation, leadership, teamwork and collaboration, communication and negotiation skills, strategic networking, career expectations, income and social status
- **at the organisational level**: Change management, workplace politics, compensations and benefits
- **at the systemic level**: Knowledge sharing, stakeholders 'interests and legitimacy

To illustrate how we mapped questions to dimension, Table 2 exemplifies the coding of questions/indicators from the TASC for the top 6 categories.

Table 2: **Examples of questions for selected dimensions**

Dimension	Question
Statistical production processes	Periodicity for collection of major surveys is defined and followed
Strategic planning and monitoring and evaluation	NSO has strategic multi-year plan updated annually which identifies organisational challenges, activities, and goals
Quality assurance and codes of conduct	NSO uses a national standard for place names and place codes for the geographic hierarchy of the country that encompasses all administrative and statistical areas and is implemented across all geospatial products
Communication	The media are given specific attention as an important outlet for statistical information to the public at large and are regularly consulted to discuss relevance, content and presentation of statistical releases
HR management	NSO has an on-the-job training programme that is supported financially and employees are given time to attend training
Human resources	Number of staff in the NSS (NSI and in statistical units of the sector Ministry) involved in production of sector statistics

3. How current assessments measure statistical capacity

a) Overall results

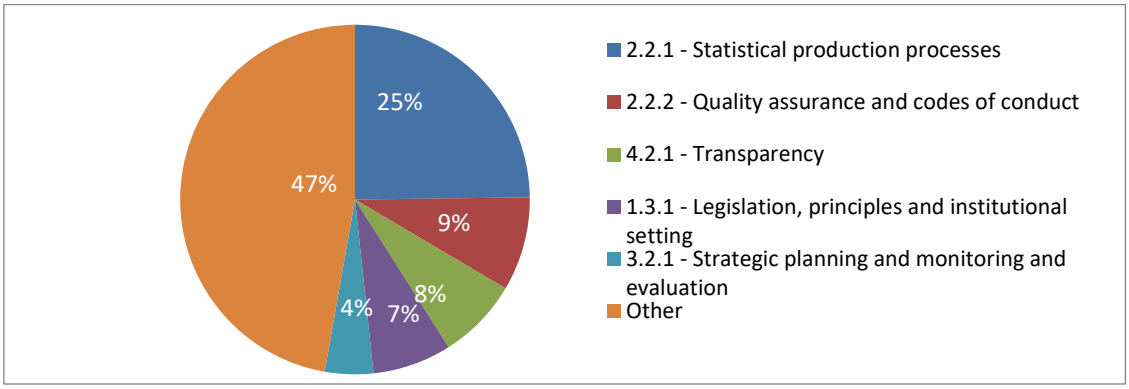
As Table 3 highlights, most of the questions across these 14 assessments focus on organisational skills and knowledge (38% of questions – three standard deviations above the mean). The organisational level absorbs 62% of overall questions, while 36% apply to the system level. Within these two levels, incentives are almost entirely neglected. Organisational politics and power, and systemic skills and knowledge, are also overlooked. There are only a few questions about individuals working for official statistical agencies.

Table 3**: Relative frequency distribution of categories, targets and levels**

	Individual	Organisational	System	Total
1 - Resources	0%	7%	16%	23%
2 - Skills and knowledge	1%	38%	1%	40%
3 - Management	0%	9%	9%	18%
4 - Politics and power	0%	7%	9%	16%
5 - Incentives	1%	1%	1%	3%
Total	2%	62%	36%	100%

Altogether, the top five dimensions represent more than half of the questions/indicators (Figure 2). Statistical production processes, quality assurance and codes of conduct account for 34% of the sample (both belong to organisational skills and knowledge). Another three (transparency, legislation, principles and institutional setting, and strategic planning and monitoring and evaluation) make up another 19%. The remaining 46 dimensions account for 47% of the sample, out of them, four contribute 0% (career expectations, teamwork and collaboration, time management and prioritisation and workplace politics).

Figure 2**: Relative frequency distribution of dimensions**



b) Results by classification

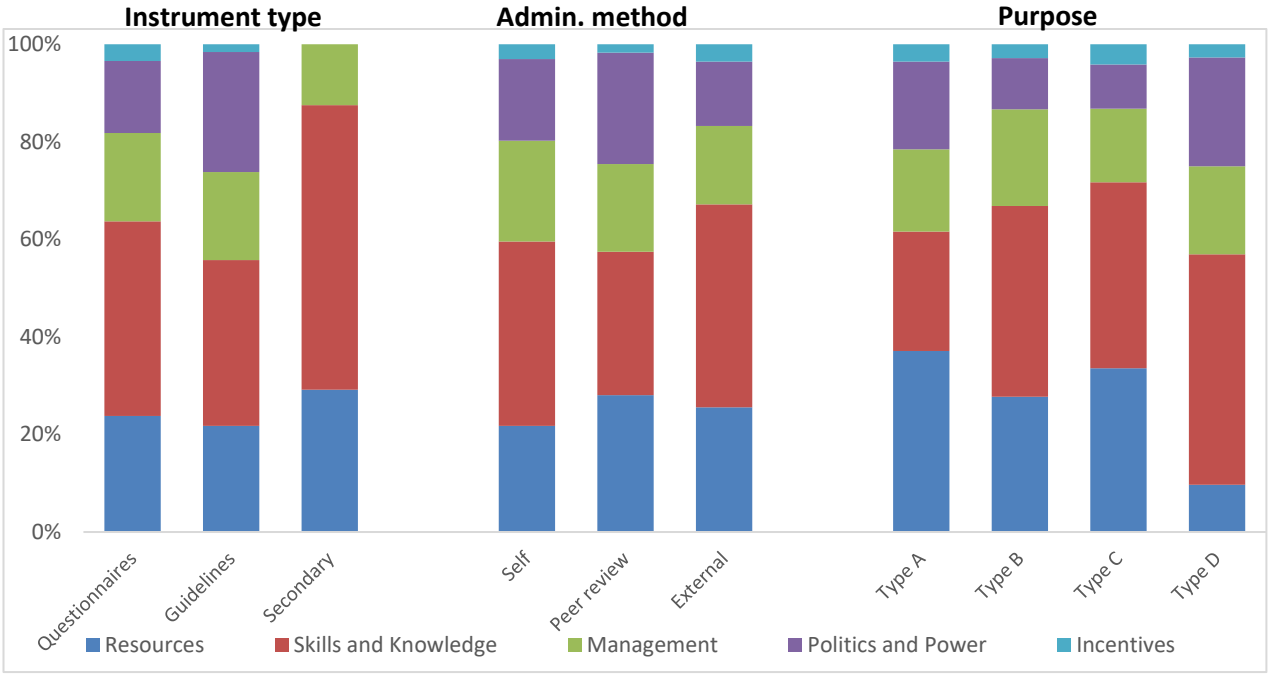
Another analysis was done on the three groups classified in Section 2d (type of instrument, administration method and purpose). The distribution of assessments in the **type of instrument** classification is uneven: ten are structured questionnaires, three are open guidelines and only one involves secondary data collection, which can only capture outputs and reporting. This focus obscures the more difficult-to-measure aspects of politics and power and incentives – thus the bulk of questions falls into skills and knowledge (Figure 3).

Questionnaires and open guidelines have greater flexibility to capture the least visible aspects of capacity. Incentives, politics and power together are addressed in less than 30% of the questions/indicators in these two groups of assessments. Yet, when testing for statistical significance in the difference between questionnaires and open guidelines using the chi-squared test, we obtained $p = 0.00$. This implies that there is a significant difference between them, with guidelines focusing more on politics, power and incentives than questionnaires.

We also found differences among the **administration methods** used for the instruments ($p = 0.00$). When inquiring into these differences, we found that peer reviews focus more on politics and power than the other methods. External reviews allotted more questions/indicators than expected to resources.

Finally, we compared types of assessments using the **purpose** classification provided by PARIS21 (described in Section 2d). Figure 3 shows the distribution of targets (i.e. resources, skills and knowledge, management, politics and power and incentives) across types. We found statistically significant differences between them ($p=0.00$). In particular, Type D assessments focus much less on resources than expected, and much more on skills and knowledge, which is consistent with their purpose of assessing output quality. They also inquire more about politics and power than expected.

Figure 3: **Comparison of distribution of targets across classifications**

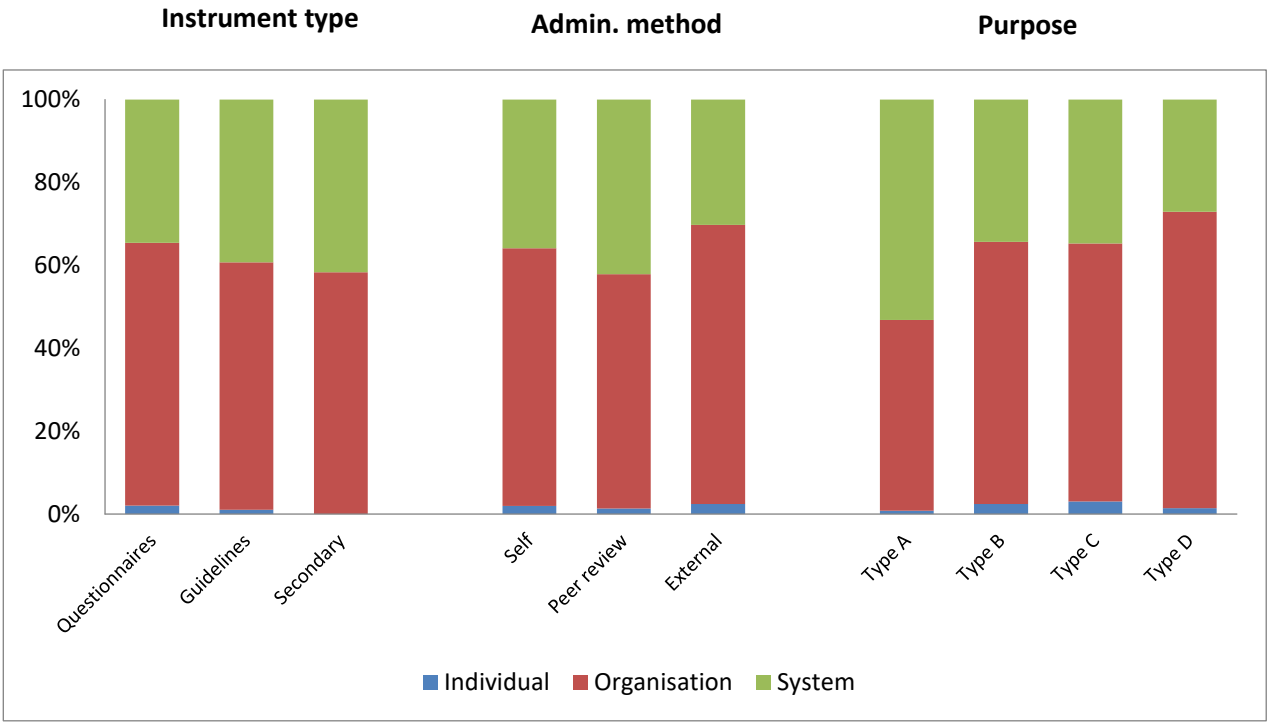


When comparing the distribution of questions across levels of capacity (Figure 4), we found that the majority of the indicators/questions regardless of **types** are devoted to the organisational level. We tested for statistical significant differences between guidelines and questionnaires and found that there were none ($p = 0.11$)

For the **administration method** we found statistically significant difference in distribution of questions/indicators between types ($p = 0.00$). Peer reviews have a higher than expected share of questions about the system, while those involving self-assessments and external reviews focus more on the organisation.

We also found significant differences amongst the **purposes classification**. These arise from the distribution of questions between the organisational and the systemic level. Type D tools focus much more on the organisation than others, while Type A tools focus more on the system.

Figure 4: **Comparison of distribution of levels across classifications**



c) Individual results

Figure 5 illustrates the distribution of questions and indicators across categories for each of the 14 assessments. Remarkably, this shows that the general pattern is not merely an aggregate effect – it is present in several assessments. Five out of the 14 (including the three longest ones) devote the largest share of questions to skills and knowledge. All of them emphasise statistical production processes equally.

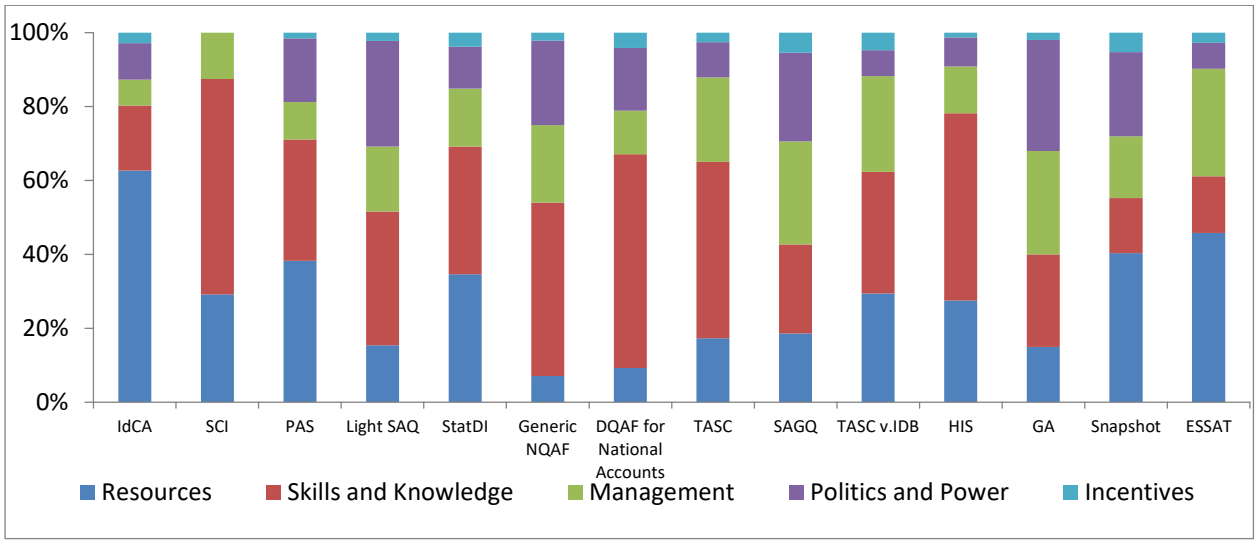
As can be noted, the IdCA, the ESSAT and Snapshot place more focus on resources. However, the IdCA places more focus on existing data (mainly assessing capacity through outputs), while the ESSAT and Snapshot stress the legislation, principles and institutional setting of the national statistical system. Finally, the GA is mainly concerned with management and politics and power – especially when it comes to strategic planning and monitoring and evaluation – as well as the transparency of the national statistical office.

Politics and power are addressed by the majority of the assessments (with the exception of the SCI); however, half of them do it in a superficial manner (less than 15% of the questions/indicators). The GA devotes 30% of its questions to this category, enquiring mainly about the relationship with users. 29% of the questions in the Light SAQ fit into this category, mostly covering transparency. On the other hand, incentives

were hardly covered by any of the reviewed assessments (though there are differences between tools – for example, SCI devotes none of its indicators to this target, while SAGQ, TASC v. IDB and Snapshot devote 5%).

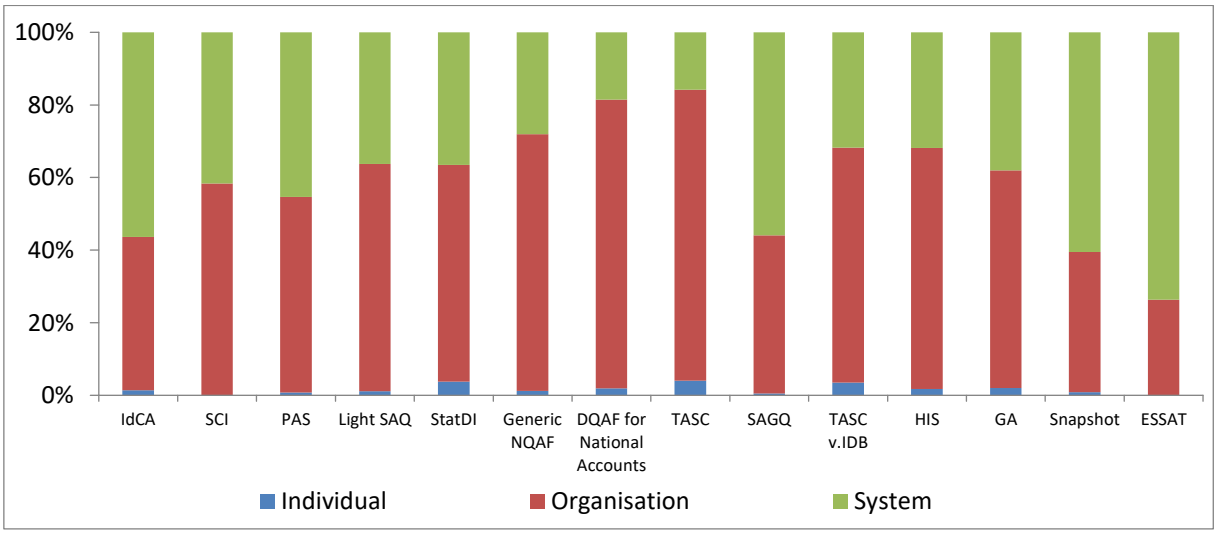
A final remark can be made regarding the modules added by the Inter-American Development Bank to the TASC, which included questions on career development of NSO employees.

Figure 5: **Comparison of category distribution between questionnaires**



When analysing the focus of questionnaires in terms of levels of capacity (Figure 6), we found that four deviate from the general pattern (whereby the main focus is on organisations), and place more focus on the system. The ESSAT, IdCA and Snapshot focus on the system resources, while the SAGQ also focuses on system management.

Figure 6. **Comparison of levels distribution between questionnaires**



d) Overlooked or under-tackled dimensions

We agree that the relative importance of dimensions may vary between agencies, and that this could imply differences in how these are covered by their assessments. We also agree that not all dimensions need the same number of questions to be correctly covered. For the purpose of identifying those dimensions that are not tackled (exhaustively) by existing assessments, we assume that those with less than 2.17% of the existing questions are under-tackled.

Table 4 shows that 31 dimensions of the framework are currently not tackled 'fairly' by existing assessments. While some would be difficult to capture, others are relatively simple – yet hardly any assessment inquires about them. For example, the reputation of a statistical organisation is difficult to capture by asking its members, unless they have conducted a survey on the topic. However, inquiring about innovation practices is simpler and very relevant to the core business of a statistical organisation.

Table 4**: Under assessed CD4.0 dimensions (those with question frequency of < 2.17%)**

	Individual		Organisations		System	
Resources	Professional background	0.33%	Human resources	2.13%	Funds infrastructure	1.38%
			Budget	0.88%		
Skills and knowledge	Technical skills	0.33%	Innovation	0.67%	Data literacy	0.75%
	Problem solving and creative thinking	0.21%			Knowledge sharing	0.50%
	Work 'know-how'	0.17%				
Management	Leadership	0.04%	Organisational design	1.21%	Advocacy strategy	0.46%
	Time management and prioritisation	0.00%	Fundraising strategies	0.33%		
			Change management	0.21%		
Politics and power	Strategic networking	0.04%	Workplace politics	0.00%	Relationship with political authorities	1.96%
	Communication and negotiation skills	0.04%			Relationship with data providers	1.13%
	Teamwork and collaboration	0.00%			Relationship between producers	0.46%
Incentives	Work ethic and self-motivation	0.46%	Reputation	0.71%	Political support	0.63%
	Income and social status	0.21%	Organisational culture	0.42%	Legitimacy	0.62%
	Career expectations	0.00%	Compensation and benefits	0.04%	Stakeholders' interests	0.13%

e) Statistical capacity in current assessments

After comparing the various purposes, types of instruments and administration modalities, we concluded that assessments portray a similar characterisation of statistical capacity, with minor differences in distribution between levels and targets. This implicit definition conveys that producing statistics is a matter of having acquired the adequate or required resources and skills at the organisation level – such as adhering to certain international standards – or having the required resources at the system level at any given point in time – such as statistical laws. The underlying assumption is that organisations and statistical systems accumulate a stock of knowledge and resources that can be mobilised for producing new statistics.

Paraphrasing Ngaruko (2008), these assessments are not able to reflect actual statistical capacity changes over time. They take a static photography of an institution at fixed points in time, which does not allow an understanding whether changes have occurred, and if they have, why. They mainly allow for flagging the existence of an issue by signalling that an activity has not been performed or an output was not produced. They do not allow for monitoring and capturing the relations between individual and organisational-level changes, system changes or outputs from capacity development – elements which are essential in order to identify capacity development (USAID, 2015).

Capacity development is a non-linear change process (FAO, n.d.): capacity can be evenly acquired and 'lost', and at least some of its aspects are in constant flux as individuals, the organisation and their environment constantly shift (USAID, 2017). Taking a series of static pictures of past or present status of capacity, with its multi-level dimensions (incentives, inputs, processes, outputs), may provide unclear perspectives into the dynamics of capacity.

Forward planning and future projections are thus an essential capacity that still needs to be measured in assessments, for instance through questions/indicators covering an organisation's future objectives, forecasted future issues etc. However, after analysing a sub-random sample of 193 questions, we found only one relating to future plans of the organisation. Nevertheless simply measuring the "potential" or likelihood of developing capacity can be misleading (USAID, 2015), especially if this potential capability focuses only on a small number of dimensions of the framework: for instance the potential for a higher budget for the NSO might be lowered in the future by the election of a government that does not support official statistics.

4. Diving deeper into selected dimensions

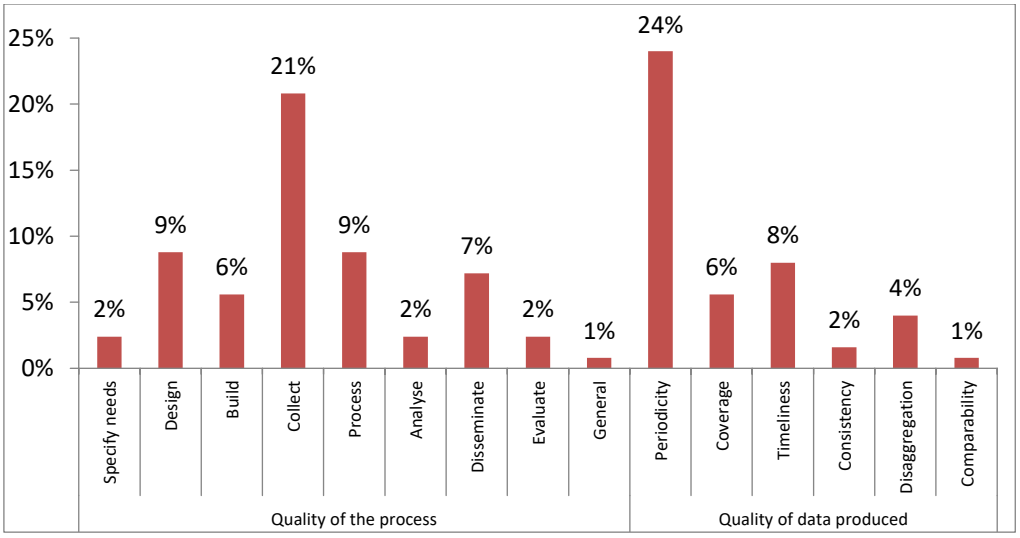
This section expands on the characterisation of capacity that we provided in the previous section, by further investigating the dimensions of our framework. The first subsection on 'Statistical production processes' expands on the content of the questions within the dimension to identify whether the large number of questions assigned to the category in fact provides a detailed coverage of the aspects included within it. The second subsection proposes questions and indicators to measure those categories that are not being tackled by existing assessments, as indicated in section 2.d. It will focus on those that receive less than 1% of the questions.

a) Organisational skills and knowledge: "Statistical production processes"

"Statistical production processes" are considered to be a particularly significant dimension of capacity. This area receives the most attention of all dimensions: 25% of the questions in the assessment tools. To provide some clarity in what these questions measure, we mapped them to the Generic Statistical Business Process Model (GSBPM) (UNECE, 2018).

We coded a random sub-sample of 125 questions (time constraint) into two sub-dimensions: the quality of the data produced and the quality of the processes associated, with the purpose of having a better understanding of what they measure. Quality is here defined as "the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs" (ISO 8402, 1994 quoted in OECD, 2006).Out of the 15 possible quality attributes and GSBPM phases that apply to these sub-dimensions, two make up nearly half of the selected questions: periodicity and collect (Figure 7). There may be several reasons for this uneven distribution: the designers of the questionnaires may have assessed the complexity of an individual process and quality attributes and determined this to be the appropriate distribution of questions; or alternatively some other aspects may be covered in 'quality assurance and codes of conduct' and other dimensions. Sixty-eight coded questions were coded into GSBPM phases and we found that almost half of them refer to the design phase. Other quality attributes or GSBPM phases may be spread across the rest of the dimensions – something that could be explored in a future analysis. What is clear, however, is that currently neither the GSPBM phases nor the attributes of data quality are evenly covered in the dimension that is most directly linked to them, ie "Statistical production processes".

Figure 7**: Relative frequency distribution of sub-dimensions**



Notes: The questions were selected from seven assessments that were conceived for assessing statistical capacity and have been structured as questionnaires: TASC, idCA, TASC v, IDB, HIS, StatDI, Snapshot and SAGQ. To measure the quality of data produced, quality attributes covered by the questionnaires include: periodicity, coverage, timeliness, consistency, disaggregation and comparability. To measure quality of the process, phases from the Generic Statistical Business Process Model (GSBPM) were used, including: specify needs, design, build, collect, process, analyse, disseminate and evaluate

b) Questions and indicators for "orphan" categories

This section suggests some new questions and indicators that could be added to assessments to improve the coverage of those dimensions of Statistical Capacity 4.0 that are currently not being tackled by the assessments we reviewed: mainly those that receive less than 1% of questions (see Table 4 in Section 3d). A few assessments do have a few questions / indicators on some of these "orphan" categories. This can be checked directly on the OAR on-line platform. We do not repeat these existing measures below, but propose complementary ones.

i. Individual resources

When assessing individual resources, indicators could cover the degree to which potential candidates or existing employees of the statistical organisation are prepared to perform their tasks without additional training, as well as the existence of a pool of skilled candidates from which to recruit. This is relevant for an organisation because it saves time and resources that otherwise would need to be invested in assembling an adequate workforce. This can be assessed indirectly by considering the labour market in general or directly by surveying employees or candidates.

Professional background

We propose the following could be considered:

- Average years of experience in the field of statistics (only for statisticians)
- Average number of different organisations worked in by staff members
- Satisfaction of the organisation's human resources' staff with the qualifications of recent graduates of statistical programmes
- Number and percentage of staff by educational attainment/qualification level? Percentage of staff indicating that their educational level is adequate for them to perform their current assigned functions well?
- Percentage of staff indicating that their education level is adequate to support further career development.
- Percentage of staff indicating that their work experience is relevant and contributing to their performance of their current assigned functions.
- Percentage of staff indicating that their work experience is adequate to support further career development.
- Percentage of staff indicating that the work experience component of the career development programme in the NSO is appropriate for their needs and aspirations.

ii. Individual skills and knowledge

Indicators related to individual skills and knowledge could capture the efficiency of the employees working for an organisation, i.e. the degree to which they produce outputs in the least amount of time and consuming the least resources. Measuring this property of the workforce from an input perspective can help to flag up the root causes of general organisational inefficiency. This can be proxied by the ability, whether formally or informally acquired, of employees to comply with their tasks or projects with minimum supervision and without disrupting the general workflow, as well as their ability to contribute to the general endeavour of the organisation. The members of the task team considered this category to be the most relevant to include in statistical capacity assessments.

Technical skills

We propose the following could be considered:

- Managers' perceptions of the adequacy of their teams' technical skills and knowledge for performing their current tasks

- Percentage of certified employees who participated in training courses specific to their current tasks
- Percentage of staff indicating that their technical skill(s) and knowledge are adequate for them to perform their current assigned functions well
- Percentage of staff indicating that their technical skill(s) and knowledge are adequate to support further career development.
- Number of contributions made by staff (statistics-related papers) to work-related professional meetings or workshops/training events, either national or international.

Work know-how

- Average number of years that employees have been working for the organisation
- Average number of departments in which employees have worked within the organisation
- Perceptions of staff members of their ability to navigate through the organisation in order to get the approvals required for their projects or tasks

Problem solving and creative thinking

- Percentage of employees reporting being micro-managed
- Managers' perceptions of staff's willingness to contribute to outlining projects and taking the lead
- Percentage of staff that have received citations for outstanding contributions to work and initiatives taken
- Satisfaction of managers with their subordinates' ability to solve problems on their own and provide innovative views on their tasks
- Percentage of employees who indicate that they regularly face challenging tasks in their assigned functions
- Percentage of staff indicating that problem solving and creative thinking are valued in the organisation

iii. Individual management

Indicators related to individual management could capture the effectiveness with which the employees conduct their tasks and also guide others in theirs, i.e. whether they conduct their activities in a way that helps to accomplish the organisation's purposes.

Time management and prioritisation

- Percentage of employees who have attended mentoring and/or training to learn how to manage their time (where such training exists in the organisation)

- Percentage of staff indicating that they schedule and prioritise their tasks
- Satisfaction of managers with their subordinates' ability to meet deadlines on time

Leadership

We propose the following could be considered:

- Percentage of high-level posts filled through promotion, rather than recruitment
- Whether the annual work plan was discussed, agreed and endorsed on time by the management of the organisation
- Frequency/number of staff meetings held by staff in management positions
- Satisfaction of the users of statistics with the leadership in the organisation

iv. Individual politics and power

Indicators in this category could cover the contribution of individuals to organisational purposes in their formal and informal exchanges with others – including staff with managers and within the management – as well as their ability to mobilise external networks to achieve strategic goals.

Teamwork and collaboration

- Percentage of staff reporting that teamwork and collaboration are valued and encouraged in the organisation

Communication and negotiation skills

- Employees' perceptions of the timeliness with which important decisions were communicated to them in the last year
- Employees' perceptions of the fluidity of communications between management and personnel
- Employees' perceptions of peers' ability to provide constructive feedback
- Employees' perceptions of the honesty and openness of their immediate supervisor

Strategic networking

We propose the following could be considered:

- Amount of job-rotations by an average employee in the past three years, including short-term exchanges with other departments in the organisation
- Percentage of staff who have subscribed to online professional networks or portals, such as Research Gate and StataList, to promote their research or exchange with other scholars in the field

- Perception of management on their ability to influence others outside the organisation, whether via formal or informal contacts, to achieve strategic goals
- Number of work-related professional meetings, workshops/training events, either national or international, attended by staff.
- Average number of professionals with whom a staff member interacts on a regular basis outside the NSS

v. Individual incentives

Incentives, though hard to capture, are relevant for the overall productivity of an organisation. Indicators in the category of individual incentives should measure employees' satisfaction with the organisation they are working for and the tasks they perform. It therefore captures their potential willingness to contribute to organisational goals for an extended period of time.

Career expectations

- Proportion of staff who consider that the organisation will provide them with career opportunities or a rewarding career
- Proportion of staff who are satisfied with the career development programme in the organisation or the national statistical system as a whole (if it is integrated)

Income and social status

We propose the following could be considered:

- Proportion of staff who report being satisfied with their income in comparison to others with equal qualifications in their profession, whether in the public or the private sector
- Proportion of staff who report being satisfied with their prestige in comparison to others with equal qualifications in their profession, whether in the public or the private sector
- Average rating given by staff in response to the invitation: "Please rate the overall income of statisticians in relation to other professions in the government, with similar qualifications" (Equal, Higher, Lower)
- Average rating given by staff in response to the invitation: "Please rate the status of statisticians in relation to other professions in the government, with similar qualifications" (Equal, Higher, Lower)

Work ethic and self-motivation

We propose the following could be considered:

- Rate of absenteeism in the past year
- Staff turnover in the past year

- Number of sanctions (following the organisations' code of conduct) in the past year
- Percentage of employees reporting satisfaction with their tasks
- Perceptions of managers of employees' interest in receiving feedback on their tasks
- Average grading given by supervisors to their staff on work ethic and motivation
- Percent of staff with an "Excellent" rating in staff performance evaluations

vi. Organisational politics and power

Indicators in this category could seek to show how power – i.e. the distribution of authority and resources – flows between individuals in the organisation, and to assess how far such relationships contribute to a nurturing work environment. The transparency dimension – which is already covered by existing assessments – measures this indirectly, since the more aligned the distribution of power with (formal) organisational objectives and codes, the more straightforward it is to inform citizens and justify decisions.

Workplace politics

- Number of years the current Director General has been in office
- Number of Director Generals in the past ten years
- Shifts in the distribution of budget between departments/units in the past three years
- Proportion of middle-managers who report that important decisions in the organisation have been made in informal settings without consulting them
- The percentage of teams and committees that successfully completed their terms of reference and within the specified time frames

vii. Organisational incentives

Indicators in this category should seek to understand what incentives the statistical organisation offers to its employees, particularly for performing at their full potential and for becoming engaged with organisational goals; and to its stakeholders, particularly for trusting the statistics it disseminates.

Compensation and benefits

We propose the following could be considered:

- Availability of bonuses for exceptional performance
- Percentage of staff on a temporary contract
- Whether all staff members (regardless of their contract situation) are offered paid leave for holidays,

sickness, pregnancy/parental leave (if applicable in the country)

- Whether staff members are offered any non-monetary benefits such as free or reduced price meals, meal vouchers, etc.

Organisational culture

We propose the following could be considered:

- Possibility for employees to demand flexible work arrangements
- Existence of areas in the office where employees can interact freely with others, such as lounges, work cafes, meeting rooms, etc.
- Awareness of employees and stakeholders of organisation's mission, vision and values
- Existence of performance reviews based on outputs and achievements
- Existence of a system of awards or recognition for innovative contributions
- Whether manager reviews include employee engagement as a metric
- Existence of a Code of Ethics for the institution
- Staff turnover rate
- Existence of formal mechanisms to ensure horizontal work between staff members
- Percentage of staff rating the institution as "Up-to-date", "Progressive", "Keeping up with changes" or "Ahead of the curve" in a staff survey.

Reputation

We propose the following could be considered:

- Percentage of users giving a high rating in response to: "Please rate the quality of leadership you observe in the NSO"
- The number of outreaches to the population at large by the NSO about statistics (brochures/leaflets, appearances on television, articles in newspapers, seminars etc.)
- Number of supportive statements (endorsements) about statistics and the statistical institutions by government representatives
- Positive rating of the NSO in a survey of users and the general public on the issues of confidence and trust

viii. System skills and knowledge

Indicators in this category could show how knowledge is shared and promoted within the national statistical system and with the actors of the data ecosystem, as well as the extent to which stakeholders (from both the NSS and external) share a common body of knowledge.

Data literacy

We propose the following could be considered :

- Recent trend (up, down, stable) in references to data and statistics in the major media.
- The number of and trend in references to data and statistics in government policy publications (development plan etc.).
- The number of website accesses and downloads
- The number of outreaches to the population at large by the NSS about statistics (brochures/leaflets, appearances on television, articles in newspapers, participation in seminars, user groups etc.)

Knowledge sharing

We propose the following could be considered:

- Rating of the extent to which staff have benefitted from senior employees transferring their knowledge and work experience in a mentoring role
- Rating of the extent to which staff have benefitted from senior employees transferring their knowledge and work experience in on-the-job training
- Rating of the effectiveness of training programmes in the career development programme (rating by staff and by "experts")
- Existence of horizontal programmes and conferences in the national statistical system
- Existence of a system or method in place for sharing knowledge among institutions of the national statistical system
- Whether all the institutions of the national statistical system participate in international task teams or working groups related to their field of expertise

ix. System incentives

Indicators in this category could measure the extent to which the national statistical system arouses interest and trust by stakeholders.

Stakeholders' interest

We propose the following could be considered:

- Composition of national working groups and committees on statistics
- Frequency with which stakeholders outside the government approach the authorities of statistical agencies
- Major interventions from the executive power in the production of headline indicators in the past period
- Number of data collection exercises or indicators that were compiled on the request of international stakeholders in the past three years

Political support

We propose the following could be considered:

- Number of supportive statements (endorsements) about statistics and the statistical institutions by government representatives in the mainstream media
- Recent trend in domestic funding of national statistical plan
- Support by government officials for requests to development partners for technical assistance and funding.

Legitimacy

We propose the following could be considered:

- Positive rating of official statistics in a survey of users and the general public on the issues of confidence and trust
- Number of supportive statements (endorsements) about statistics and the statistical institutions by government representatives

5. Conclusion

Our work aims to stimulate the discussion around how statistical capacity is, and should be, measured. By assessing 14 of the most popular international statistical capacity assessments against the criteria of the PARIS21 Capacity Development 4.0 framework, we found that existing assessments mainly focus on identifying whether statistical organisations and systems have acquired the adequate or required resources and skills at any given point in time. This depiction of capacity gives the incorrect perception that there is a linear progression from being 'incapable' to being 'fully capable' of producing statistics. The assessments mostly overlook the individual and systemic dimensions of the CD4.0 frameworks.

To further advance the measurement of statistical capacity, consistently with the CD4.0 framework, we propose new indicators for categories not currently covered.

This work could be further complemented in many different ways. Two priorities could be explored. First, measuring the response burden on countries, to know how many times the same information was requested from them. This would require collecting additional information on which assessments were implemented and where. This would involve direct input from the international agencies behind each assessment, as this information is generally not publicly available. Second, the analysis could be complemented by performing a refined qualitative content analysis to detect similarities, differences and duplications between questions/indicators. This could help to harmonise approaches and questions and would enhance the way we collectively measure statistical capacity, keeping in mind the well understood specificities and priorities of each agency.

Annex 1: List of Questionnaires

Assessment	Abbreviation	Organisation	Instrument type	Assessment administration	Purpose classification
In-Depth Country Assessment	idCA	FAO	Questionnaire	External	National planning
Snapshot	Snapshot	EUROSTAT	Questionnaire	Self and External	Statistical programme design and monitoring
African Statistical Development Indicators	StatDI	UNECA	Questionnaire	Self	Global Monitoring of statistical performance
Tool for Assessing Statistical Capacity	TASC	US CENSUS BUREAU	Questionnaire	Self and External	Statistical programme design and monitoring
Assessing the National Health Information System. An Assessment Tool	HIS	HMN/WHO	Questionnaire	Self	Statistical programme design and monitoring
Generic National Quality Assessment Framework	Generic NQAF	UNSD	Questionnaire	Self	Data quality assurance and compliance with codes of practice, norms and standards
Self-Assessment Guidance Questionnaire	SAGQ	UNECA	Questionnaire	Self	National Planning
Data Quality Assessment Framework for National Accounts Statistics	DQAF for National Accounts	IMF	Questionnaire	External	Data quality assurance and compliance with codes of practice, norms and

					standards
Global Assessment of the National Statistical System	GA	EUROSTAT/UNECE/EFTA	Open guidelines	Self and Peer	Data quality assurance and compliance with codes of practice, norms and standards
Tool for Assessing Statistical Capacity	TASC v. IDB	IDB	Questionnaire	Self and External?	Statistical programme design and monitoring
Environment Statistics Self-Assessment Tool (Part I Institutional Dimension of Environment Statistics)	ESSAT	UNSD	Questionnaire	Self	Statistical programme design and monitoring
Light Self - Assessment Questionnaire on the implementation of the European Statistics Code of Practice	Ligt SAQ	EC/EUROSTAT	Open guidelines	Self	Data quality assurance and compliance with codes of practice, norms and standards
Pan-African Statistics Programme: Peer Reviews of NSIS/NSSS in African countries	PAS	EUROSTAT/AUSTAT	Open guidelines	Peer	National planning
Statistical Capacity Indicator	SCI	WB	Secondary data collection	External	Global monitoring of statistical performance

References

Denney, L. and Mallett, R. with Benson, M. S. (2017), "Service delivery and state capacity: findings from the Secure Livelihoods Research Consortium", Secure Livelihoods Research Consortium, London, https://securelivelihoods.org/publication/service-delivery-and-state-capacity-findings-from-the-secure-livelihoods-research-consortium

Eurostat (n.d.), The Eurostat Quality Assurance Framework, Eurostat, Brussels, https://unstats.un.org/unsd/dnss/docs-nqaf/Eurostat_QAF%20leaflet.pdf.

Ngaruko, F. (2008), "The World Bank's framework for statistical capacity measurement: strengths, weaknesses, and options for improvement", *The African Statistical Journal*, 7.

OECD (2006) Quality-ISO in OECD Glossary of Statistical Terms. Retrieved March 29th, 2018 from: https://stats.oecd.org/glossary/detail.asp?ID=5150

PARIS21 (2018a), Proposing a Framework for Capacity Development 4.0 (draft), PARIS21, Paris, www.paris21.org/results-cd40-task-team

Strode, M. (2017), *Guide to Assessment of National Statistical Systems*, PARIS21, Paris.

UNECE (2018) Generic Statistical Business Process Model. In *UNECE Statistics Wikis.* Retrieved March 29th, 2018 from: https://statswiki.unece.org/display/GSBPM/Generic+Statistical+Business+Process+Model

World Bank (2017), *Data on Statistical Capacity*, The World Bank, Washington DC, http://datatopics.worldbank.org/statisticalcapacity, accessed 22 November 2017.

PARIS21

WWW.PARIS21.ORG