Assuring quality in the new data ecosystem: Mind the gap between data and statistics!

Matthias Reister[1,2]
United Nations Statistics Division, New York, NY, USA
E-mail: reister@un.org

Abstract. Drawing on recent work to develop the United Nations National Quality Assurance Frameworks Manual for Official Statistics to respond to the new data ecosystem, this paper addresses three important questions now facing the statistical community: (1) How can official statistics assure the quality of data from administrative and other sources? (2) Can the quality assurance framework for official statistics be applied to data as opposed to statistics? (3) What other implications does the difference between data and statistics have for the role of official statistics in the new data ecosystem? The paper argues that statistical offices should strongly support the establishment of national data stewards but should not take on such a role themselves. Mixing responsibilities for data and official statistics risks both undermining official statistics and not doing justice to the need to develop data as an asset in a responsible way.

Keywords: Data and statistics, data ecosystem, data stewardship, official statistics, quality assurance

1. Recognizing the impact of the new data ecosystem on official statistics

The release of the 'Data Revolution' report, 'A World That Counts' in November 2014 by the United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development [1] marked a milestone and turning point for the development of official statistics.[3] It clearly articulated the need to foster innovation to fill data gaps by bringing together traditional and new data sources (including Big Data) and creating new infrastructures for data development and sharing, while developing global ethical, legal and statistical standards to improve data quality and protect people from abuses in a rapidly changing data ecosystem.

The report and subsequent discussions have highlighted the unprecedented level of growth in data and how many aspects of our lives are being captured in data. Specifically, new technologies allow a previously unimaginable level of recording, analysis, and integration of data about human behaviors and overall societal trends. Data is the key component of many business models, and many enterprises are managing and utilizing very large amounts of data, including data that can be linked to individuals. Aware of the importance of data and the risks of misuse, many countries, including the United Kingdom [3], the United States of America [4], China [5] and the European Union [6] have been adopting national data strategies and have been passing legislation to regulate the collection, storage, and use of data on national and transboundary levels.

The emergence of this new data ecosystem, which can be defined as a system in which several actors interact to exchange, produce, and utilize data [2, p. 5], presents both opportunities and challenges for official

---

[1]The views expressed herein are those of the author and do not necessarily reflect the views of the United Nations.

[2]The paper greatly benefited from comments by and discussions with Hans Viggo Saeboe, Norway. The author is also very grateful for the review, comments, and suggestions by Simon Scott and two anonymous reviewers.

[3]Official statistics describe, on a representative basis, economic, demographic, social and environmental phenomena of public interest. Official statistics are developed, produced and disseminated as a public good by the members of the national statistical system in compliance with the Fundamental Principles of Official Statistics and accepted quality frameworks as well as other internationally agreed statistical standards and recommendations [2, pp. 6–7].

statistics. New data sources can improve the efficiency or timeliness of official statistics by using already available data. Statistical offices may produce less data themselves and may become more often curators of official statistics produced by others. At the same time, the role of the national statistical office (NSO)[4] and other producers of official statistics as the predominant providers of data and statistics for policy makers and the public may be challenged, as users are able to resort to data and statistics produced outside the national statistical system or even by themselves. Statistical authorities have been responding to this challenge by stressing the value of official statistics as a trusted source of quality information and the core values of official statistics [7], by modernizing their statistical production processes and use of data, and by expanding the provision of data, analysis, and information. At the same time, many leaders of NSOs are rethinking the overall role of official statistics in the new data ecosystem and discussing their organizations' possible future roles in national data stewardship.[5]

2. How the UN manual on quality assurance for official statistics responds to the new data ecosystem

At its fiftieth session in March 2019, the United Nations (UN) Statistical Commission adopted the UN National Quality Assurance Frameworks Manual for Official Statistics (Manual) [2]. The Manual updated and replaced a template and guidelines issued in 2012. Its recommendations on statistical quality assurance, its updated statistical quality assurance framework and its implementation guidelines all aim to address the requirements of the new data ecosystem.

2.1. Two core recommendations addressing the scope of quality assurance of official statistics

The Manual stipulates that all members of the national statistical system should apply the national quality assurance framework to all data and statistics used for government decision-making. Specifically,

1. "It is recommended that countries establish a national quality assurance framework for official statistics and that all members of the national sta-

tistical system commit to continually assessing, improving and reporting on the quality of official statistics, as well as on the quality of data and statistics used in the production of official statistics as required" (Recommendation 3).

2. "It is recommended that the national quality assurance framework be implemented at the national statistical office and throughout the entire national statistical system. Furthermore, it is recommended that the national quality assurance framework be applied to all data and statistics produced outside of the national statistical system that are disseminated with the help and support of a member of the national statistical system or that are used for government decision-making, as deemed appropriate and required" (Recommendation 5) [2, para. 2.6].

2.2. Important updates to the UN quality assurance framework

Chapter 3 and the Annex of the Manual update the UN National Quality Assurance Framework (UN-NQAF) from the earlier version of 2012 to reflect the need to actively utilize new data sources:

1. Under Principle 1 'Coordinating the national statistical system', requirement 1.3 demands that "there is a mechanism for considering statistics produced outside the national statistical system, and if appropriate, for these statistics to become official."

2. Under Principle 2 'Managing relationships with data users, data providers and other stakeholders' three requirements (2.5–2.7) address the need to gain adequate access to and utilize administrative data and data (including "Big Data") maintained by private corporations or other non-governmental organizations for statistical purposes on a regular basis, including for testing and experimentation.[6] The NSO should also cooperate with and provide support and guidance to data providers.

3. Principles related to the management of statistical processes call for promoting innovation in the development, production and dissemination of statistics (requirement 10.5), efforts to improve the statistical potential of administrative data and other data sources (requirement 12.3) and for data

---

[4]Here taken to have the same meaning as national statistical institute.

[5]The 67th plenary session of the Conference of European Statisticians in Paris on 26–28 June 2019 may be viewed as the start of the wider discussion on data stewardship within the official statistics community that has continued ever since.

---

[6]Other quality assurance frameworks for official statistics such as the European Statistics Code of Practice also address the need for data access.

sharing, data linkage and use of administrative and other data sources to minimize the respondent burden (requirement 13.4).

4. Under Principle 14 'Assuring relevance', requirement 14.3 is that "statistics based on new and existing data sources are being developed in response to society's emerging information needs".

2.3. *The Manual's discussion of the use of different data sources*

The Manual contains several chapters on the implementation of a national quality assurance framework throughout the national statistical system. Specifically, Chapter 7 addresses quality assurance when different data sources are used to produce official statistics. It distinguishes between statistical data sources, administrative data sources and other data sources according to their purpose and the entity responsible for data compilation and discusses the use of multiple data sources. The Manual defines these three different data sources as follows:

1. Statistical data sources are data collections created primarily for official statistical purposes by government agencies or other entities working on behalf of the government. Statistical data sources include statistical sample surveys, censuses, and statistical registers. Statistical data sources are often referred to as primary data sources while administrative and other sources are secondary sources.

2. Administrative data sources are data sets created primarily for administrative purposes by government agencies or other entities working on behalf of the government. Administrative data sources include administrative registers of persons and legal entities and the records of ministries, departments, and specialized agencies, such as tax returns, social services records and customs data, or data of regional or local administrations.

3. Other data sources include all data sets that are not created primarily for official statistical or administrative purposes but rather for commercial or other private purposes. They include data sets created by providers of communications, media and e-commerce services, providers of services based on Earth observation and remote sensing, and private insurance companies, but also through traditional sample surveys conducted by companies for their own purposes, such as market research [2, para. 7.4].

The Manual takes the view that the UN-NQAF and other generic national quality assurance frameworks apply to the production of official statistics regardless of the data source. However, it is recognized that the challenges to obtain compliance can be different depending on the data source [2, para. 7.1].

3. Approaches to assess the quality of source data to produce official statistics

There is a strong and increasing need to assess the quality of administrative and other data sources to produce official statistics. However, assessing the quality of source data is part of, but also different from assessing the quality of a statistical product. One can see the assessment of source data as a separate step that may warrant special attention when using administrative and other data sources.

3.1. *Typical quality challenges in the use of administrative and other data sources*

The Manual identifies numerous challenges in using administrative or other data sources to produce official statistics, including access and co-ordination problems, lack of proper use of statistical concepts, lack of transparency, the need to assure appropriate statistical procedures and respect for principles of confidentiality, relevance, accuracy and reliability (especially as regards coverage), comparability and metadata. For example, telecommunication data may indicate movements of people, but we will often not know how representative they are of the total population and hence the coverage of the data, and comparability problems may arise from changing patterns in the use of telecommunication services. At the core of such challenges is the fact that administrative and other data sources are not geared towards the production of official statistics. Statistical agencies have only limited or no influence at all on what data is being compiled, unlike in the case of statistical data sources that are specifically oriented towards the production of official statistics. Chapter 7 of the Manual provides a list of examples of specific elements to be assured when statistical, administrative, other, or multiple sources of data are used.

3.2. *Approaches to assess the quality of source data*

This section briefly introduces several approaches to assessing the quality of source data with a view to using

them to produce official statistics.[7] These approaches may also be relevant for the assessment of data quality in general – a concern often raised by data users and producers in government, civil society and the private sector.

The European BLUE-ETS project provided an overview of measurement methods to assess the quality of administrative data when used as an input for official statistics [8]. Its indicators were grouped under five dimensions of quality: technical checks (technical usability of the file and its data), accuracy (data are correct, reliable and certified), completeness (under-coverage or over-coverage), time dimensions (including timeliness and punctuality) and integrability (extent to which the source data can be linked up with administrative and other data). In addition, the project identified preconditions for use such as legal access, public acceptance, availability of a unified identification system, comprehensive and reliable systems of public administration and cooperation among authorities that must be met.

Following up on earlier work [9], the United Nations Economic Commission for Europe published guidelines for assessing the quality of administrative data for use in censuses [10]. These consider quality at four stages: at the source, at the stage when data is received by the NSO, at the process stage, and at the output stage. The quality dimensions at the source comprise relevance and accuracy (for use in the census), timeliness, coherence and comparability, accessibility (the ease with which the NSO can obtain the data) and interpretability (availability of metadata). The institutional environment of the data holder (including its capacity) is also included. The quality dimensions at the data-receiving stage are validation and harmonization (e.g., a readable file format), accuracy and reliability (for variables and population coverage), timeliness and punctuality, and linkability (e.g., a common unique identifier). Together, the source and data receiving stages give an assessment of input quality.

Other important efforts are the development of checklists for the evaluation of the quality of input data [11] as part of the Eurostat program ESS Vision 2020 AD-MIN [12]. These checklists identify a consolidated list of six quality dimensions and 17 associated indicators. Also, already in 2014, the UNECE Big Data Quality

Task Team suggested a Framework for the Quality of Big Data [13] which uses a hierarchical structure composed of the three hyper-dimensions: source, metadata, and data. More recently, De Broe et al. have discussed quality criteria for integrating new data and methods in official statistics [14] and concluded that the European Statistical System quality principles as laid down in the European Statistics Code of Practice and hence also in the UN-NQAF, do not need to be adapted to the emergence of new data sources. De Broe et al. argue that new quality aspects associated with big data sources can be integrated into existing quality frameworks at the level below the respective quality principles.

3.3. Conclusions for further work

In conclusion, important work has been undertaken to guide assessments of the quality of source data to produce official statistics. There are commonalities among these approaches such as the use of the quality principles of official statistics associated with statistical outputs and a focus on characteristics of the dataset. However, at this time, there is no single or uniform answer on how to assure the quality of data from administrative and other sources to produce official statistics. Maybe this is unavoidable, but there would be benefit in having some list of criteria and/or indicators and guidelines that would provide at least a common starting or reference point for everyone concerned with the quality of source data from administrative or other data sources.

4. Why quality assurance frameworks for official statistics cannot be directly applied to data

4.1. The distinction between data and statistics

Frequently, there are calls to address the "quality of data and statistics" and to only use "quality data" for decision-making. This paper argues that it is very important and useful to make a distinction between data and statistics when discussing quality assurance of official statistics. Often the terms "data" and "statistics" are used synonymously, or statistics is viewed as part of the larger set of data. The Manual is more precise. It defines statistics as numerical information relating to an aggregate of data on units or observations. It uses the term "statistics" when referring to an output of a statistical production process and the term "data" when referring to input or possibly throughput in that process. The term "microdata" is a special case as depending

---

[7]The section is partly adapted from the work of Hans Viggo Saeboe in support of the United Nations Expert Group on National Quality Assurance Frameworks and presented at a side event to the 53rd session of the United Nations Statistical Commission on 10 February 2022.

on the context it can also be an output [2, p. 5]. The term "data" encompasses not only the data produced and used for statistical purposes, but the infinitely larger universe of data produced by the public and private sectors through the use of technology for many different purposes and potentially available for many uses.

4.2. Quality assurance is driven by user needs

Let us recall the standard definition of quality. Quality is the degree to which a set of inherent characteristics of an object fulfils requirements [15]. A simple definition of quality is "fit for use" or "fit for purpose". It is the users' needs that define quality. User needs can differ and must be reconciled.

The UN-NQAF is geared towards the quality assurance of official statistics – the aggregates which are the end-product of the statistical production process. The user needs associated with official statistics are largely encapsuled by Principle 1 of the Fundamental Principles of Official Statistics (FPOS) [16], which speaks of the need to satisfy citizens' entitlement to public information concerning the economic, demographic, social and environmental situation. All official statistics are produced for the purpose of providing users with specific statistical information. Quality assurance frameworks for official statistics such as UN-NQAF assess the quality of input data or source data only from the perspective of the ultimate purpose of producing specific statistical outputs to satisfy user needs.

The user needs associated with data are much broader and all data is *a priori* multi-use. Data can support any conceivable action and interaction of humans including the production, exchange and consumption of private and public goods and services, independently of whether the data was initially intended for these purposes. Therefore, requirements or fitness for use of data for those different purposes can be very different from the requirements to produce official statistics.

4.3. Quality assurance frameworks for official statistics cannot be directly applied to data

Given the different nature of user needs, statistical quality assurance frameworks such as UN-NQAF are not geared towards assessing and improving the quality of data used for non-statistical purposes. They cannot be directly applied to data in general. This is illustrated in the following examples:

1. Many of the requirements reflected in the UN-NQAF principles 14–18 concerning relevance, accuracy and reliability, timeliness and punctuality, accessibility and clarity, and coherence and comparability of statistical outputs are also relevant for data. However, they must be adapted or applied differently depending on what the data is being used for. For example, the principle of relevance applied to official statistics addresses the question whether statistical outputs meet the current information needs of users. In contrast, the needs of users of *data* can be different, widely varying and largely unknown to us as statisticians. We may know some but not all users. Therefore, it can be difficult or impossible to identify and balance user needs for data, whereas the uses and users of official statistics are typically already defined in the design of the statistical production process. Another example is the requirement for official statistics to conduct revision studies to assure accuracy and reliability. It is clear what revision studies mean for statistical outputs, but it is much less clear what they could mean for data, which in many cases is only generated once.

2. Similarly, some of the general requirements and best practices for managing statistical processes are also applicable to data. For example, UN-NQAF principle 10 points to the need to evaluate source data and to improve methods and promote innovation to ensure methodological soundness, and principle 12 recognizes the need to facilitate data linkages and to use unique identifiers for statistical units. However, many of the requirements of official statistics such as the use of international standards or statistical processes have no, or no clear, applicability to data.

3. One of the main principles of official statistics (and one of its core values) is statistical confidentiality, based on Principle 6 of FPOS which states that data is only to be used for statistical purposes. In practice, statistical laws may grant access to individual-level data for research purposes if specific conditions are met. By contrast, the objective associated with data as reflected in the different national data strategies is not to restrict and prevent but rather to encourage different uses while at the same time preventing misuse. Often, it is the information about individuals or individual statistical units that makes data useful and valuable. Hence, statistical confidentiality can be and is often detrimental to the use of data.

4.4. Conclusions for further work beyond official statistics

While quality assurance frameworks for official statistics cannot be directly applied to data, the effort

M. Reister / Assuring quality in the new data ecosystem: Mind the gap between data and statistics!

to assess the quality of source data to produce official statistics as described in Section 3 could form the basis for analyzing the quality requirements of data in general. However, such work would need to start with researching what is already being done in the private and public sectors and be undertaken in cooperation with respective holders and users of data. In this context, it is important to mention that the United Kingdom (UK) has already developed a Government Data Quality Framework [17] not focused on official statistics but on data to support the government's ambitions related to the digital transformation of public services and the UK becoming a world leader in the use of artificial intelligence. The framework is based on the following five principles: 1. Commit to data quality, 2. Know your users and their needs, 3. Assess quality throughout the data lifecycle, 4. Communicate data quality clearly and effectively, and 5. Anticipate changes affecting data quality. The UK's approach contrasts notably with an OECD report on the path to becoming a data-driven public sector, which takes the fitness for purpose of data largely for granted and focuses instead on data governance and data management [18]. These divergent perspectives indicate, at least in the view of this author, an urgent need to deepen and advance the discussion of data quality assurance in the public sector.

5. The implications of the difference between data and statistics for the discussion of data stewardship and other issues

As shown above, data and statistics are different, and quality assurance frameworks for official statistics cannot be directly applied to data in general, even if some quality criteria apply to both. These two conclusions have strong implications for the current discussions in official statistical circles on data stewardship. The issues they raise are of central importance because the quality assurance frameworks for official statistics encapsulate their purpose and core values. Careful consideration of the differences between data and statistics can also bring greater clarity to several other important discussions affecting official statistics.

5.1. Current discussions on data stewardship and other topics conflate data and statistics

Numerous recent discussions bear on the issue of NSOs' stewardship role in the new data ecosystem. The Conference of European Statisticians concluded that

NSOs' roles are changing in response to the ecosystem's new demands and opportunities, accelerated by the COVID-19 crisis. It points out that NSOs can position themselves in the new data ecosystem in different ways, ranging from minimal change to full data stewardship [19]. The UN Working Group on Data Stewardship came to a similar conclusion, noting that there are different interpretations of data stewardship and that its definition and application must be context-specific and will depend on the possible roles of NSOs. The Working Group points out that effective data stewardship aims to increase trust in data, and its value, use and impact [20]. An earlier note by Estonia (with contributions of others) to the Conference of European Statisticians suggested that data stewardship should, among other things, include supporting high quality and optimized use of data [21].

The discussions of data stewardship at the UN Statistical Commission and the Conference of European Statisticians generally do not make an explicit distinction between statistics and data. The note by Estonia on the new role of NSOs speaks about the role of NSOs and official statistics while referring at the same time to enabling the use of data. Similarly, the report of the UN Working Group refers to "data and statistics" or only to "data" when discussing the role of the NSOs in data stewardship.

The tendency to speak of data and statistics in one breath, without distinguishing between them, is not limited to the discussion of data stewardship. The newly adopted terms of reference of the UN Statistical Commission (Resolution E/RES/2022/3) refer to the Commission's responsibility for statistical and data-related systems [22]. Also, the discussion of the integration of geospatial and statistical information frequently does not distinguish the terms statistics and data. The same applies to the discussions on open data within the statistical community or when international donor funding for "data and statistics" is discussed. In some cases, the conflation may be justified, but it is worth considering several areas in which it may be important for NSOs to carefully observe the implications of the differences between data and statistics.

5.2. Implications of the distinction between data and statistics for the discussion of data stewardship

The note by Estonia to the Conference of European Statisticians in June 2020 [21] provided a first list of activities and skills that should be included under data stewardship within public data governance. These in-

clude supporting high quality and optimized use of data, facilitating access to data, promoting expertise, ethics, skills and data literacy, promoting common standards, frameworks and data policies, and elaborating data strategies. Further work under the auspices of the Conference, and ongoing work under the UN Statistical Commission, have broadened and deepened the discussion of data stewardship, focusing particularly on data governance and issues of trust, equity, and inclusion.

However, no agreed definition and list of functions of a national data steward has yet emerged, probably due to the wide variety of countries' approaches to data stewardship. In the absence of an agreed stipulative list, we may for the purposes of our discussion use a normative definition of the basic function of a data steward, without going into the details of how this would be implemented in different country situations. Thus, based on the stated objectives of various national data strategies, the primary function of a data steward would be to foster the use of data derived from either governmental or private sources by anybody and for any purpose unless it violates existing laws and regulations.

Based on this concept of data stewardship, conscious of the distinction between statistics and data, and recognizing that quality assurance frameworks embody official statistics' purpose and core values, this paper offers the following preliminary conclusions.

1. *The role of data steward poses an inherent conflict for an NSO.* The NSO is responsible for the production of official statistics according to the FPOS to satisfy specific information needs. Its core values and operations as reflected in the national quality assurance frameworks for official statistics are geared towards this task and not to the task of fostering the use of data by anyone for any purpose. This conflict is well illustrated by Principle 6 of FPOS. As already mentioned, this states that individual data collected by statistical agencies for statistical compilation are strictly confidential and used exclusively for statistical purposes. Taken literally, this means that all data collected by the NSO through statistical surveys, census, and registers and maybe also all other data compiled by the NSO from administrative and other sources can only be used for statistical purposes. In practice, it is accepted that statistical laws define exceptions from this principle such as for research purposes or the sharing with other producers of official statistics under certain restrictive conditions [23]. However, this is clearly not sufficient for a data steward tasked to make data available for all acceptable uses. NSOs may already be mitigating the effect of confidentiality restrictions by providing micro-

data sets and legislation to allow the re-use of data for other purposes. Yet taking this beyond tight limits might well undermine the trust of data providers, including respondents to statistical surveys, and/or require elaborate privacy-protection measures. The use of data will require a different approach to privacy protection from that implemented for official statistics.

2. *There are different possibilities for reconciling the roles of data steward and producer of official statistics.*

- First, there is the status quo or wait-and-see option, where a country implicitly or explicitly decides that there is as yet no need for a data steward.
- A second option is the creation of a data stewardship function at the NSO with limited scope that does not impact on the NSO's prime role as statistical agency.
- Thirdly, the role of data steward may be given to an organization other than the NSO, with the NSO in a supporting role.
- A fourth possibility is to broaden the mandate of the NSO to include key elements of data stewardship, but to keep this function and responsibility separate from the NSO's role in the production of official statistics.
- A fifth possibility is to merge the responsibilities of producing official statistics with those of data stewardship. This would fundamentally alter the basic concept of the NSO as a statistical agency and probably require the revision of national statistical laws to reconcile the requirements and principles of official statistics with the requirements of the role of data steward. At global level, this may imply the revision or replacement of the FPOS should many NSOs follow this path.

Different country circumstances may determine which of these options is feasible. However, as a general rule it seems that NSOs should not try to simply subsume the data stewardship role themselves under their existing mandates. As this paper has shown, there are tensions between the basic aims of official statistics policies and data/Big Data policies, so that merging responsibilities for them in a single body risks both undermining official statistics and not doing justice to the need to develop data as an asset in a responsible way.

3. *Data stewardship is a 'can't miss' – opportunity for official statistics.* Data access and adequate resources are key requirements to produce official statistics as reflected in the existing quality frameworks for official statistics. The existence of a data steward holds the promise of dramatically improved access to administrative and private data for everyone and probably even

more so for official statistics as it has already strong internal safeguards to protect the privacy of data. This is critical as new sources come to replace old ones. Furthermore, it can be expected that any additional public and private resources for data and statistics will primarily focus on the development and use of national data assets rather than the development of additional official statistics. Existing resources for official statistics may even be re-programmed towards data, as seems to be happening already in the case of support for statistical development.[8] Also, it will probably be more efficient to establish data centers that allow the use of AI by all of government than to make separate arrangements for official statistics. Therefore, it appears critical for NSOs to support the establishment of a national data stewardship function as it would allow better access to data and help overcome resource constraints such as those related to costly surveys or the use of new technologies such as AI. Also, NSOs may use the establishment of a national data steward as an opportunity to further define their own role in the new data ecosystem, whatever this role may be.

5.3. Implications of the distinction between data and statistics for the discussion of other important issues in official statistics

Making the distinction between data and statistics also offers additional clarity on other issues:

1. There has been intensive work on open data within the statistical community. The importance of the open data principles as articulated in the International Open Data Charter [25] for official statistics cannot be overstated. However, there are many similarities between the principles of the Charter and the statistical quality principles of UN-NQAF, and most official statistics are already open. Hence, the discussion of open data is much more important for data than for official statistics and should primarily pursued in fora for data.
2. There are many discussions and guidelines on the integration of geospatial and statistical information. Some clearly distinguish statistics and data, others do not. It is important to note that the natural type of integration of both is on the data level, i.e., the integration of geospatial and statistical data on the input or production level of official statistics. Statistical outputs are aggregates that have a geospatial dimension, but the value of integration at that level is often more limited.
3. There is a need to clearly distinguish funding for data and funding for official statistics. For example, there have been over the years many calls to increase international donor support for "data and statistics" and the information typically presented on funding levels does not make a distinction between support for official statistics and data systems [26]. This is in certain respects understandable, giving the close link between certain administrative systems such as for civil registration and vital statistics with the production of official statistics. However, if the national statistical system of a developing country is to be held accountable for its ability to provide official statistics such as on the SDG indicators, then there is a need for information about the support it receives. Also, as the importance of data increases, a shift of funding away from official statistics towards data may go unnoticed unless funding for data is clearly distinguished from funding for official statistics. Maintaining this distinction may also offer better insights into the effectiveness of support. Also, as it concerns domestic resources, increased total funding for data and statistics may initially benefit official statistics. However, given the attractiveness and dynamic development of the use of data, official statistics may become over time only one of many functions of data. Both are different and important in their own right and therefore need their own funding.

6. Summary of findings

The UN National Quality Assurance Frameworks Manual for Official Statistics makes a clear distinction between data as input and statistics as output of the production process of official statistics. Careful observation of this distinction leads to the conclusion that quality assurance frameworks for official statistics cannot be directly applied to data in general, or only to a limited degree. There have been important efforts to systematize assessment of the quality of administrative and other sources to produce official statistics, but these vary significantly in their approaches. There may be a

---

[8]Official development assistance (ODA) to developing countries for projects with a primary focus on data and statistics has largely stagnated over the last 10 years despite hugely increased requirements for SDG monitoring and despite an overall increase of ODA. At the same time, the number and funding of projects with a data and statistics component has dramatically increased [24].

benefit of having a common list of criteria or indicators and basic guidelines that would provide at least a uniform starting point or reference for everyone concerned with the quality of source data to produce official statistics. This could also assist in the identification of quality criteria for data in general, which would need to be developed in cooperation with governmental and private producers and users of data.

Distinguishing data from statistics can help bring clarity to the discussion of the role of official statistics in the new data ecosystem. In particular, the distinction is very important for the discussion of data stewardship and whether the NSO should take on such role. This paper argues that the dual roles of custodian of official statistics and the role of national data steward cannot be easily reconciled. At the same time, NSOs should support the establishment of a national data steward as it is likely to facilitate the production of official statistics.[9] Paying closer attention to the differences between data and official statistics should also help to scope out other aspects of the future of official statistics in the new data ecosystem, including questions relating to open data, integration of geospatial information, and support for statistical development.

References

[1] United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG). A World That Counts: Mobilizing The Data Revolution for Sustainable Development. 2014. Available at: https://www.undatarevolution.org/report/?msclkid=b6010481ade211eca5f451c29211e84f [last accessed January 26, 2023].

[2] United Nations. United Nations National Quality Assurance Frameworks Manual for Official Statistics, Series M No. 100, New York. 2019. Available at: https://unstats.un.org/unsd/methodology/dataquality/un-nqaf-manual/ [last accessed January 26, 2023].

[3] United Kingdom Government. National Data Strategy. 2020. Available at: https://www.gov.uk/government/publications/uk-national-data-strategy [last accessed January 26, 2023].

[4] US Government. Federal Data Strategy. 2019. Available at: https://strategy.data.gov/ [last accessed January 26, 2023].

[5] People's Republic of China. Personal Information Protection Law. 2021. Available at: https://personalinformationprotectionlaw.com/ [last accessed January 26, 2023].

[6] European Union. General Data Protection Regulation (GDPR). Available at: https://gdpr.eu/tag/gdpr/ [last accessed January 26, 2023].

[7] Task Team on Core Values. Core values of official statistics, at Conference of European Statisticians, Seventieth plenary session, Geneva (ECE/CES/2022/2). 2022. Available at: https://unece.org/sites/default/files/2022-07/ECE_CES_ 2022_ 2-2211176E.pdf [last accessed January 26, 2023].

[8] Daas P, Ossen S. BLUE-ETS: Deliverable 4.2: *Report on methods preferred for the quality indicators of administrative data sources*. 2011. Available at: http://www.pietdaas.nl/beta/pubs/pubs/BLUE-ETS_WP4_Del2.pdf [last accessed January 26, 2023].

[9] United Nations Economic Commission of Europe. Guidelines on the use of registers and administrative data for population and housing censuses, Geneva. 2018. Available at: https://unece.org/guidelines-use-registers-and-administrative-data-population-and-housing-censuses-0# [last accessed January 26, 2023].

[10] United Nations Economic Commission of Europe. Guidelines for assessing the quality of administrative sources for use in censuses, Geneva. 2021. Available at: https://unece.org/statistics/publications/CensusAdminQuality [last accessed January 26, 2023].

[11] Eurostat. ESSnet KOMUSO, Work package 1: Checklist for Evaluation the Quality of Input Data. 2016. Available at: https://ec.europa.eu/eurostat/cros/content/wp3-quality_en [last accessed January 26, 2023].

[12] Eurostat. ESS Vision 2020 ADMIN (Administrative data sources). 2020. Available at: https://ec.europa.eu/eurostat/cros/content/ess-vision-2020-admin-administrative-data-sources_en [last accessed January 26, 2023]. (2020)

[13] United Nations Economic Commission for Europe Big Data Quality Task Team. A suggested Framework for the Quality of Big Data. 2014. Available at: https://statswiki.unece.org/download/attachments/108102944/Big%20Data%20Quality%20Framework%20-%20final-%20Jan08-2015.pdf?version=1&modificationDate=1420725063663&api=v2 [last accessed January 26, 2023].

[14] De Broe S, et al. Updating the Paradigm of Official Statistics: New Quality Criteria for Integrating New Data and Methods in Official Statistics. Statistical Journal of the IAOS. 2021; 37(1). 343-360.

[15] International Organization for Standardization (ISO) 9000: 2015.

[16] United Nations. Fundamental Principles of Official Statistics. Resolution 68/261 adopted by the General Assembly on January 29, 2014. A/RES/68/261. 2014. Available at: https://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf [last accessed January 26, 2023].

[17] UK Government Data Quality Hub. The Government Data Quality Framework, published 3 December 2020. 2020. Available at: https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework#foreword [last accessed January 26, 2023].

[18] OECD. The Path to Becoming a Data-Driven Public Sector, OECD Digital Government Studies, OECD Publishing, Paris. 2019. Available at: 10.1787/059814a7-en [last accessed January 26, 2023].

[19] United Nations Economic Commission for Europe. Terms of Reference for the Task Force on Data Stewardship. ECE/CES/BUR/2021/FEB/4. para. 4. 2021. Available at: https://drupal-main-staging.unece.org/sites/default/files/2021-02/04_Data%20stewardship%20TOR_appr.pdf [last accessed January 26, 2023].

[20] United Nations. Report of the Working Group on Data Stewardship. E/CN.3/2022/5, paras. 27-29. 2021. Available

---

[9]While this is not the topic of this paper, the author has the view that there is an urgent need for a national institution or function within government that fosters and addresses the use of data that is properly resourced for this task.

at: https://unstats.un.org/UNSDWebsite/statcom/session_53/
documents/2022-5-DataStewardship-E.pdf [last accessed January 26, 2023].

[21] United Nations Economic Commission for Europe. Implementation of the new role of national statistical offices at the time of expanded possibilities. ECE/CES/2020/10, para. 34. 2020. Available at: https://unece.org/fileadmin/DAM/stats/documents/ece/ces/2020/ECE_CES_2020_10-2005282_E. pdf [last accessed January 26, 2023].

[22] United Nations. Ensuring that the work in the field of statistics and data is adaptive to the changing statistical and data ecosystem. Resolution 2022/3 adopted by the Economic and Social Council on 8 June 2022. E/RES/2022/3. 2022. Available at: https://unstats.un.org/UNSDWebsite/statcom/session_53/documents/TOR-E.pdf [last accessed January 26, 2023].

[23] United Nations Economic Commission for Europe. Generic Law on Official Statistics. 2016. Available at: https://unece.org/statistics/publications/generic-law-official-statistics [last accessed January 26, 2023].

[24] PARIS21. Partner Report on Support to Statistics 2021, Paris. 2021. Available at: https://paris21.org/press2021 [last accessed January 26, 2023].

[25] Open Data Charter. International Open Data Charter. Available at: https://opendatacharter.net/principles/ [last accessed April 26, 2023].

[26] PARIS21. The PARIS21 Partner Report on Support to Statistics 2022: A Wake-Up Call to Finance Better Data, OECD Publishing, Paris. 2022. Available at: 10.1787/c3cfb353-en [last accessed January 26, 2023].