Workshop on the Implementation of a National Quality Assurance Framework for Official Statistics in countries of the Africa Region

Addis Ababa, Ethiopia, 14-18 October 2019

Session 7: Quality assurance for statistics compiled from different data sources

Matthias Reister, Chief
Development Data Section
Statistics Division, Development Data and Outreach Branch
United Nations Department of Economic and Social Affairs

Quality assurance for statistics compiled from different data sources (Chapter 7)

❑ This Manual distinguishes three data sources <u>according to their purpose and by the entity responsible for their compilation</u>.

❑ These are <u>statistical data sources</u>, <u>administrative data sources</u> and <u>other data sources</u>.

❑ Chapter 7 highlight certain aspects of quality assurance that are specific, or are of special importance, to statistics that are produced (completely or partially) using specific data sources.

Quality assurance for statistics compiled from different data sources (Chapter 7)

Statistical data sources:

❑ 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 <u>sample surveys,</u> different types of <u>censuses</u> <u>and statistical registers</u>. Statistical registers can themselves be derived from different sources (often administrative data sources);

Quality assurance for statistics compiled from different data sources (Chapter 7)

Statistical data sources:

❑ <u>The main advantage:</u> allow data to be obtained according to specified needs and pre-defined statistical concepts.

❑ <u>Major quality challenges</u>:

a. High cost of production (Principle 11 - Cost-effectiveness);

b. Availability of resources (Principle 9 - Adequacy of resources);

c. Low frequency of conducting sample surveys and censuses (Principle 14 - Relevance);

d. Respondent burden and willingness of respondents to provide information (Principle 13 – Respondent burden);

e. Sampling and non-sampling errors (Principle 15 - Accuracy and reliability);

f. Need for complex sampling designs (Principle 10 - Methodological soundness);

g. The need for careful planning, implementation of instruments, training and supervision of staff and rigorous evaluation (Principle 12 – Appropriate statistical procedures).

Quality assurance for statistics compiled from different data sources (Chapter 7)

Overview table: Examples of some specific elements to be assured when different data sources are used

UN NQAF Principle	Data Source	Requirements / Elements to be assured*	Explanation
		Managing the statistical system	
Principle 1: Coordinating the national statistical system	Statistical	There is a body that ensures coordination of sample surveys and their methodological soundness throughout the NSS. (1.2, 10.1)	Sample surveys should be coordinated and integrated for cost-effectiveness and methodological soundness
	Administrative	Administrative records are systematically linked with records from other relevant administrative data systems, as permitted by applicable laws (1.2, 10.1 and 11.5)	The linking of administrative data will allow the best possible use of the already available information
	Administrative Other	There is a unit that discusses and provides support for the use of administrative data sources and other data sources within the NSS (1.2, 10.1 and 11.5)	The use of administrative and other data sources poses unique challenges that should be addressed by sharing experiences and best practices

Quality assurance for statistics compiled from different data sources (Chapter 7)

Administrative data sources

❑**Administrative data sources** are datasets created primarily for administrative purposes by government agencies or other entities working on the behalf of the government

❑Administrative data sources include <u>administrative registers</u> of persons and legal entities and the <u>records of ministries, departments and specialized agencies</u>, such as tax returns, social services records and customs data, or data of regional or local administrations.

Quality assurance for statistics compiled from different data sources (Chapter 7)

Administrative data sources:

❑ <u>The main advantage:</u> cost-effectiveness, reduction of respondent-burden, improved timeliness, and improved relevance, accuracy and reliability due to its ability to obtain highly disaggregated data.

❑ <u>Major quality challenges</u>:

a. Insufficient cooperation with the providers/holders of data, frequent lack of clarity on the roles and responsibilities of different stakeholders, and legal challenges to obtain access (Principle 1 – Coordinating the national statistical system, and Principle 2 - Managing relationships with data users, data providers and other stakeholders);

b. Incoherent use or lack of use of statistical standard concepts, definitions and classifications (Principle 3 – Managing statistical standards);

c. Lack of explicit quality commitment of holders (producers) of administrative data (Principle 8 – Quality commitment);

d. The concepts that are underlying administrative data sources are not reflecting the statistical concepts that are used to measure the phenomena (Principle 14 – Assuring relevance);

Quality assurance for statistics compiled from different data sources (Chapter 7)

Administrative data sources:

❑ <u>Major quality challenges</u>:

e.   Methodological and technical difficulties in managing access to administrative data sources, performing record linkage and integrating data across multiple administrative data sources, transmitting the data and integrating them with data from statistical sources (Principle 12 – Appropriate statistical procedures);

f.   Preserving security and preserving confidentiality of individual data when disseminating detailed statistics may be a challenging task (Principle 7 – Statistical confidentiality and data security);

g.   Interference and bias in the statistical production process and lack of information about how the data has been produced (Principle 4 – Professional independence, Principle 5 – Impartiality and objectivity, Principle 6 – Transparency and Principle 19 – Managing metadata);

h.   Under or over-coverage of the target population; mis-alignment with the statistical reference period; inherent and unquantified bias as a result of the original purpose of the administrative dataset (Principle 15 – Assuring accuracy and reliability);

i.   Inability to quantify uncertainty when there are only non-sampling errors (Principle 15 – assuring accuracy and reliability).

Quality assurance for statistics compiled from different data sources (Chapter 7)

Other data sources

❑ **Other data sources** are all datasets that are not created primarily for official statistical or administrative purposes but rather for commercial or other private purposes

❑ 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 include traditional sample surveys conducted by companies for their own purposes, such as market research.

❑ In general, other data sources include data sources associated with the term "big data" unless already included, in some instances, in statistical or administrative data sources.

Quality assurance for statistics compiled from different data sources (Chapter 7)

❑ **List of other data sources (no classification)**

   a. Cross-country sample surveys by supra-national organizations or international enterprises;

   b. Data compiled and maintained by private professional organizations or business associations or non-profit institutions in general;

   c. Data and records compiled and maintained and/or owned by enterprises that cover large parts of the population of statistical units, in particular e-commerce. media and telecommunication providers but also other enterprises that provide services directly to individuals or businesses such as insurance companies, banks, airlines etc.;

   d. Earth observation and remote sensing;

   e. Thematic mapping and monitoring systems (e.g., field-monitoring stations for water quality, air pollution etc.);

   f. Research/scientific and pilot studies;

   g. Citizen generated data

❑ **Note: New data sources can be often associated with other data sources but may as well belong to statistical or administrative data sources, depending on national circumstances.**

Quality assurance for statistics compiled from different data sources (Chapter 7)

Other data sources:

❑ <u>The main advantage</u>: opportunity to overcome resource limitations, to allow much more frequent and timely reporting, provide more objective information and, most importantly, to be able to generate data on phenomena and its aspects (dis-aggregations) that are difficult or impossible to capture with traditional statistical and administrative data sources. This leads to improved relevance. .

❑ <u>Major quality challenges</u>:

a.   The limited access to other data sources and legal challenges regarding its access (as it can also be the case for administrative sources) requires arrangements with the data providers (e.g. government agencies, private sector and research institutions); lack of knowledge about the existence of such data; and sustainability of the source over time (Principle 2 - Managing relationships with data users, data providers and other stakeholders);

b.   Incoherent use or lack of use of statistical standard concepts, definitions and classifications (Principle 3 – Managing statistical standards) that put the accuracy, reliability, coherence and comparability of the resulting statistics in question;

Quality assurance for statistics compiled from different data sources (Chapter 7)

Other data sources:

❑ Major quality challenges:

c. Providers of data (which may be the owner or holder of the data) are not subject to and do not adhere to the Fundamental Principles of Official Statistics (FPOS) and associated statistical quality principles such as professional independence (Principle 4) and quality commitment (Principle 8);

d. Utilizing data for statistical purpose may potentially put the confidentiality and privacy of individuals, households and businesses at risk depending how detailed data is being published (Principle 7 – Statistical confidentiality and data security);

e. Data of sources such as those from mobile phones or social media are not representative of the entire population and may cause serious selection bias when used for statistical purposes (Principle 10 – Methodological soundness; Principle 12 – Appropriate statistical procedures and Principle 15 - Accuracy and reliability);

f. …

Quality assurance for statistics compiled from different data sources (Chapter 7)

Multiple data sources:

❑ <u>The main advantage</u>: increasingly common in the compilation of official statistics, especially for the monitoring of the SDG; disaggregated data may be obtained through the integration with data from a different source, e.g. combining data sources may provide a higher coverage of the target population.

❑ <u>Major quality challenges</u>:

a. Assuring methodological soundness (Principle 10) as related to different coverage and the use of different concepts and definitions etc.;

b. Assuring appropriate statistical procedures (Principle 12) as related to having appropriate procedures, skills and knowledge to be able to link and integrate data;

c. Assuring confidentiality and data security (Principle 7);

d. Assuring coherence and comparability over time (Principle 18); the quality of the entire integration process should be assessed by calculating proper indicators;

Chapter 7 contains selected references to quality assurance when specific data sources are being used.

THANK YOU