

Session 3: Quality assurance of administrative and other (new) data sources

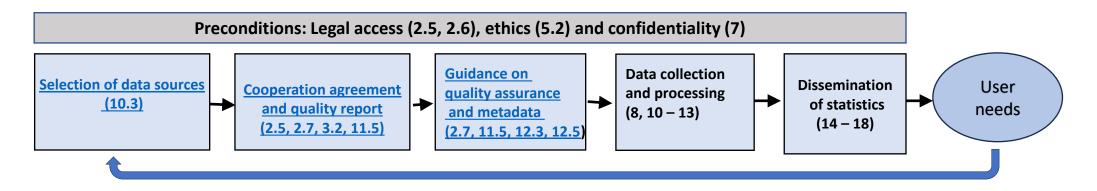
# 3.4 Discussion of implementation cases and review of available implementation guidance and options

Matthias Reister, <a href="mailto:reister@un.org">reister@un.org</a>
Chief, Development Data Section
Statistics Division, Development Data and Outreach Branch
United Nations Department of Economic and Social Affairs

Workshop on the Implementation of a National Quality Assurance Framework for Official Statistics in Countries of the Latin American and Caribbean Region

Bogota, Colombia, 22-24 November 2023

## Important steps when using administrative and other data for official statistics – and their link to UN NQAF



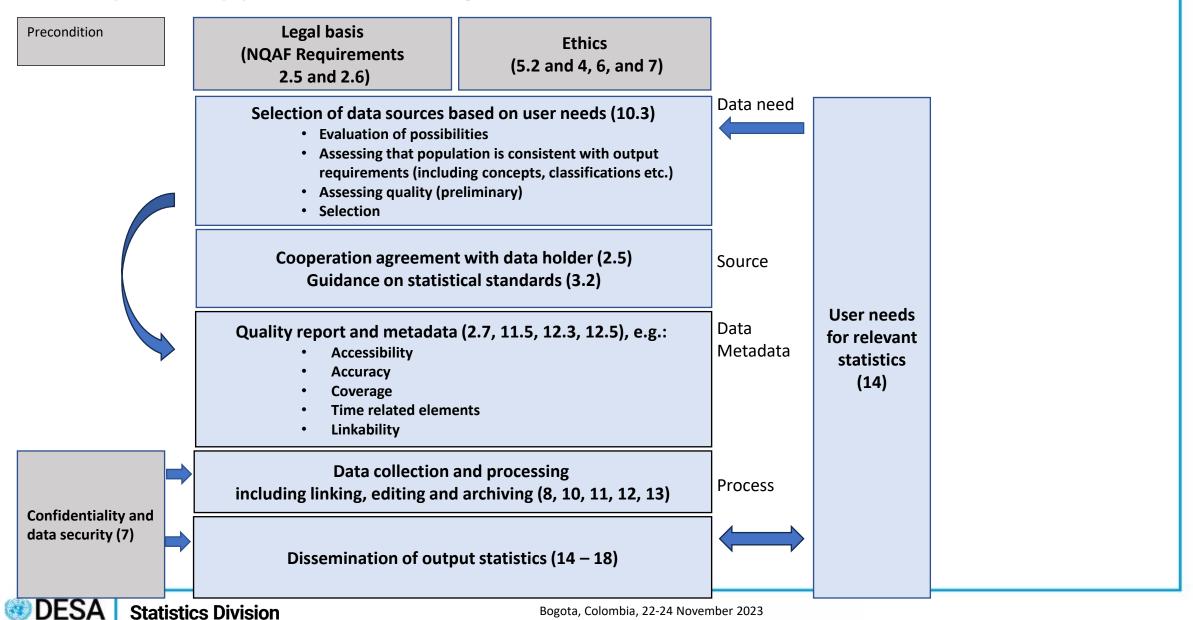
Source data is selected based on user needs

A cooperation agreement should include terms of delivery and quality requirements

Quality reports specifies the most important input data quality dimensions Guidance comprises use of standards, quality assessment and metadata.

Processes include linking, imputing, coding, editing, development of statistical registers and archiving Official statistics fit for user needs

#### A conceptual approach to using administrative and other data sources



# Question: How is the use of admin and other data sources reflected in GSBPM

Overarching Processes							
Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Reuse or build collection instruments	4.1 Create frame and select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult and confirm needs	2.2 Design variable descriptions	3.2 Reuse or build processing and analysis components	4.2 Set up collection	5.2 Classify and code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Reuse or build dissemination components	4.3 Run collection	5.3 Review and validate	6.3 Interpret and explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame and sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit and impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing and analysis	3.5 Test production systems		5.5 Derive new variables and units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare and submit business case	2.6 Design production systems and workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production systems		5.7 Calculate aggregates			
				5.8 Finalise data files			

#### What NQAF covers or does not cover

- NQAF includes most of the factors affecting quality of statistics based on administrative and other data sources, though not always at the same level of detail. The main principles/requirements/elements covered are largely summed up in the figures.
- Possible main requirements and elements from the examples that are not covered comprise:
  - Review of the data holder organization (and its reputation to ensure that it is ethically and legally acceptable to acquire the data)
  - Quality assurance of the data holder (beyond assessment of its data)
  - Details about the content of Cooperation agreements/MoUs
  - Details about the content of quality reports (a checklist on quality dimensions of input data, see next slide)
  - Details about data processing beyond linkage, editing and archiving (such as imputing, coding and development of statistical registers)
- Terminology may vary in the examples, there is a need for some harmonization



## Example 1: Quality criteria for administrative data

Statistics Netherlands 2009

- Quality criteria linked to Source: Supplier, Relevance, Privacy and security, Delivery, Procedures
- Quality criteria linked to Metadata: Clarity, Comparability, Unique keys, Data treatment (by supplier)
- Quality criteria linked to **Data:** Technical checks, Coverage, Linkability, Non-response, Measurement, Processing, Sensitivity



<u>45-Checklist-quality-evaluation-administrative-data-sources-2009.pdf</u>

### **Example 2: Quality criteria for administrative data**

- Technical checks (usability)
- Accuracy (erroneous objects or variables)
- Completeness (under-coverage, over-coverage, redundancy)
- Time (timeliness and punctuality)
- Integrability (linking-ability including comparability with variables from different data sources)

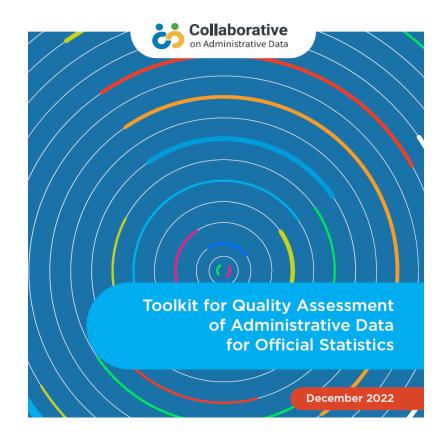
Metadata must be available

Source: **BLUE-ETS Project** 

## **Example 3: UN Collaborative on Administrative Data**

- 1. The Statistical System, Institutional Framework and Agreements
- 2. Input data quality (Source)
- 3. Quality and data processing:
  - Accuracy and completeness (including over and under coverage, selectivity, and redundancy)
  - Validity
  - Clarity (metadata)
  - Coherence and Consistency.
  - Technical checks
- 4. Ensuring quality of statistical outputs

Source: Other 04.a Toolkit for Quality Assessment of Admin Data for Official Statistics.pdf



#### Example 4: Quality criteria for other and new data: Big data

- 1. Quality criteria linked to **Source**:
  - Institutional environment (sustainability of data provider, transparency and interpretability)
- 2. Quality criteria linked to **Metadata**:
  - Complexity, completeness, usability (of metadata) etc.
  - Time

#### 3. Data quality:

- Accuracy and selectivity, including coverage (over and under-coverage and imperfections)
- Linkability etc.

Source: <u>UNECE Big Data Quality Task Team</u>

A Suggested Framework for the Quality of Big Data

Deliverables of the UNECE Big Data Quality Task Team

December, 2014

#### Roundtable 1: Possible tools / guidance – Pros and Cons

- 1. Amendment / supplement to UN NQAF
  - Countries to integrate the additional requirements into their NQAF
- 2. Module(s) on admin and other data sources (group exercise)
  - Focuses exclusively on the relevant requirements for admin and other data sources
  - Can be used in conjunction with existing NQAF
  - Complements your NQAF
- 3. Separate toolkit
  - Includes all other relevant requirements of your NQAF, such as on statistical outputs
  - No need to refer to NQAF (advantage for some users)
  - Overlaps with NQAF and risks confusion
- 4. All of the above, other suggestions?



# Roundtable 2: Do you find the conceptual approach useful? A conceptual approach to using administrative and other data sources

