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

3.3 Review of applicable principles, requirements, and elements to be assured, and their mapping to UN NQAF

Matthias Reister, reister@un.org
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
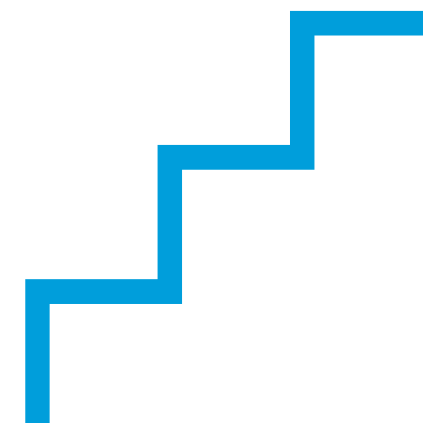
1. Mapping of practices and guidelines to UN NQAF

2. Identification of important steps when using administrative and other data for official statistics – and their link to UN NQAF

3. Identification of critical requirements

4. Identification of critical requirements - What NQAF covers or does not cover

5. Review of critical requirements when using administrative and other data sources – ROUND TABLE DISCUSSION

Examples of practices, guidelines and tools

Officially shared during 2022 global consultation
1. Netherlands, Kindom of the: Netherlands_45-Checklist-quality-evaluation-administrative-data-sources-2009.pdf
2. Australia: Australia_Quality Management of Statistical Outputs Produced From Administrative Data.pdf
3. United Kingdom: UK_Quality-Assurance-of-Administrative-Data.pdf;
UK_Quality Assurance Toolkit_updated_Feb19_2.pdf
https://best-practice-and-impact.github.io/admin-data-quality-stats/departments.html#Introduction
4. Italy: Italy_Guidelines for the quality of statistical processes that use administrative data.pdf
5. United Arab Emirates: UAE_Manual of Statistical Quality Standards and Procedures for Administrative.pdf
6. Canada: Available at https://www.statcan.gc.ca/en/data-quality-toolkit#a4 (11 pages)
7. Oman: Oman_ben_Quality of Administrative Records.pdf
8. Norway: Norway_PM2022-08.pdf
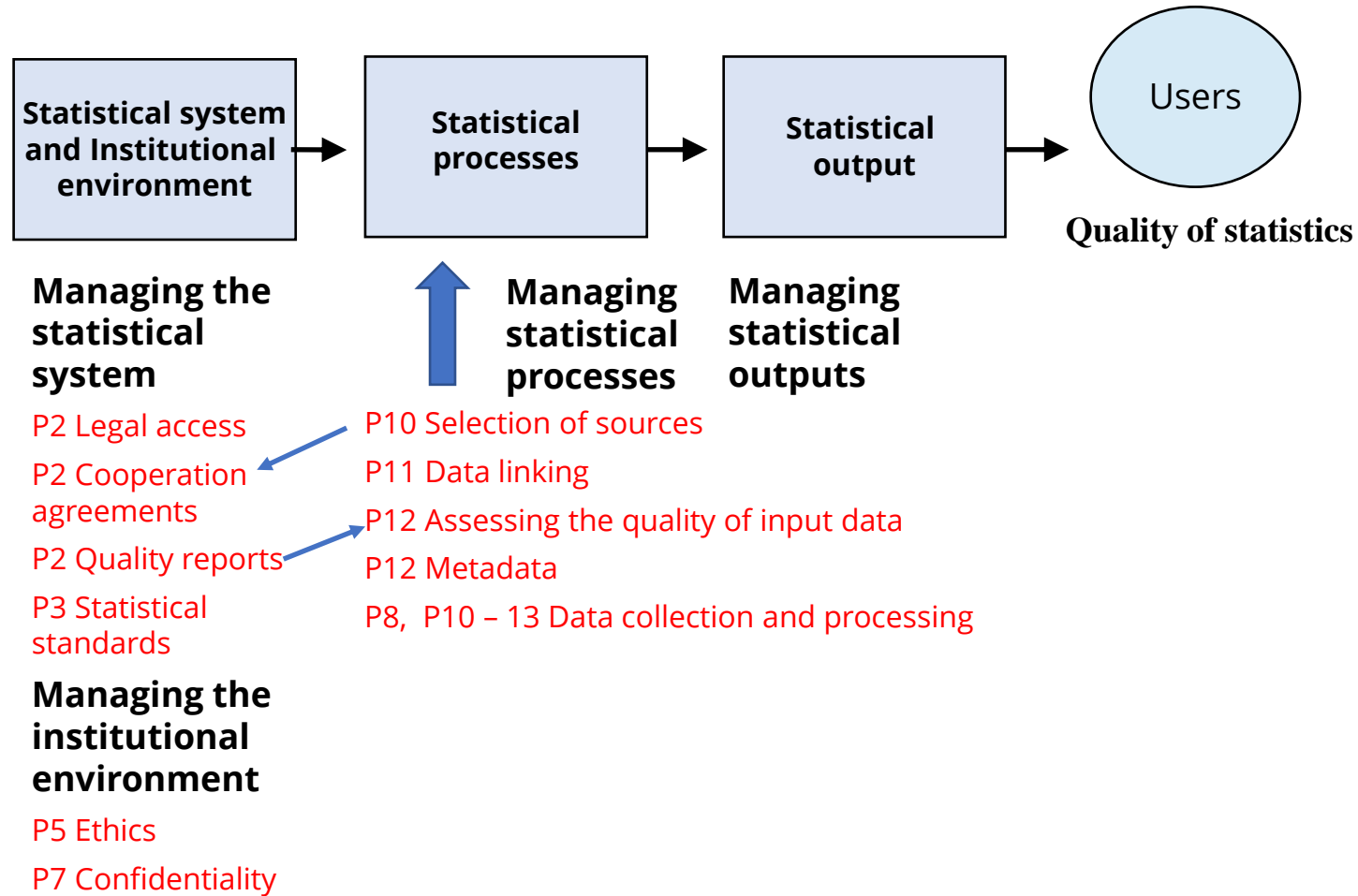9. Poland: Poland_Tools, metrics and quality reporting of administrative data sources.docx

Other
1. BLUE-Enterprise and Trade Statistics (BLUE-ETS): BLUE-ETS_WP4_Del2.pdf
2. UNECE - Guidelines for Assessing the Quality of Administrative Sources for Use in Censuses: ECECESSTAT20214_WEB - UNECE 2021 Guidelines.pdf
3. ESSnet KOMUSO Quality in Multisource Statistics - Work Package 1 Checklist for Evaluating the Quality of Input Data: Other 03. essnet_wp1_report_final_version4 (KOMUSO).pdf
4. Collaborative on Administrative Data - Toolkit for Quality Assessment of Administrative Data for Official Statistics
Other 04.a Toolkit for Quality Assessment of Admin Data for Official Statistics.pdf
Other 04.b Questionnaire for assessing quality of admin data.xlsx
5. Data Quality Assessment Tool for Administrative Data, Iwig, W., et al., 2013: Other 05.a US DataQualityAssessmentTool.pdf
6. ESSnet (use of admin and accounts data), WP6, Quality Indicators when using Administrative Data in Statistical Outputs: https://cros-legacy.ec.europa.eu/system/files/SGA%202011_Deliverable_6.1.pdf
7. Guidelines for Quality Assessment of Administrative Data, National Institute of Statistics of Rwanda, 2018: Other 07. Guidelines for Quality Assessment of admin. data-R.docx
8. Checklist used by Colombia: Other 08. GPD-051-PDT-001-f-001_Documento_Revisión_Calidad_Mayo2023.xlsx
9. Self-assessment tool for admin data used by ILO (an adaptation for labour statistics of the CECRA questionnaire recommended by the Working Group on Administrative Records in the framework of the Statistical Conference of the Americas (SCA) and coordinated by ECLAC): a) Other 09.a SAQUAR Users Guide_EN (Ver.2).docx b) Other 09.b Questionnaire SAQUAR_EN (Ver.2).docx;
10. A Framework for Data Quality, USA, 2020, Other 10. FCSM.20.04_A_Framework_for_Data_Quality.pdf
11. Revised Version of the Quality Guidelines for the Acquisition and Usage of Big Data, ESSnet, 2020, Other 11. WP3_Deliverable_K3_Revised_Version_of_the_Quality_Guideline.pdf
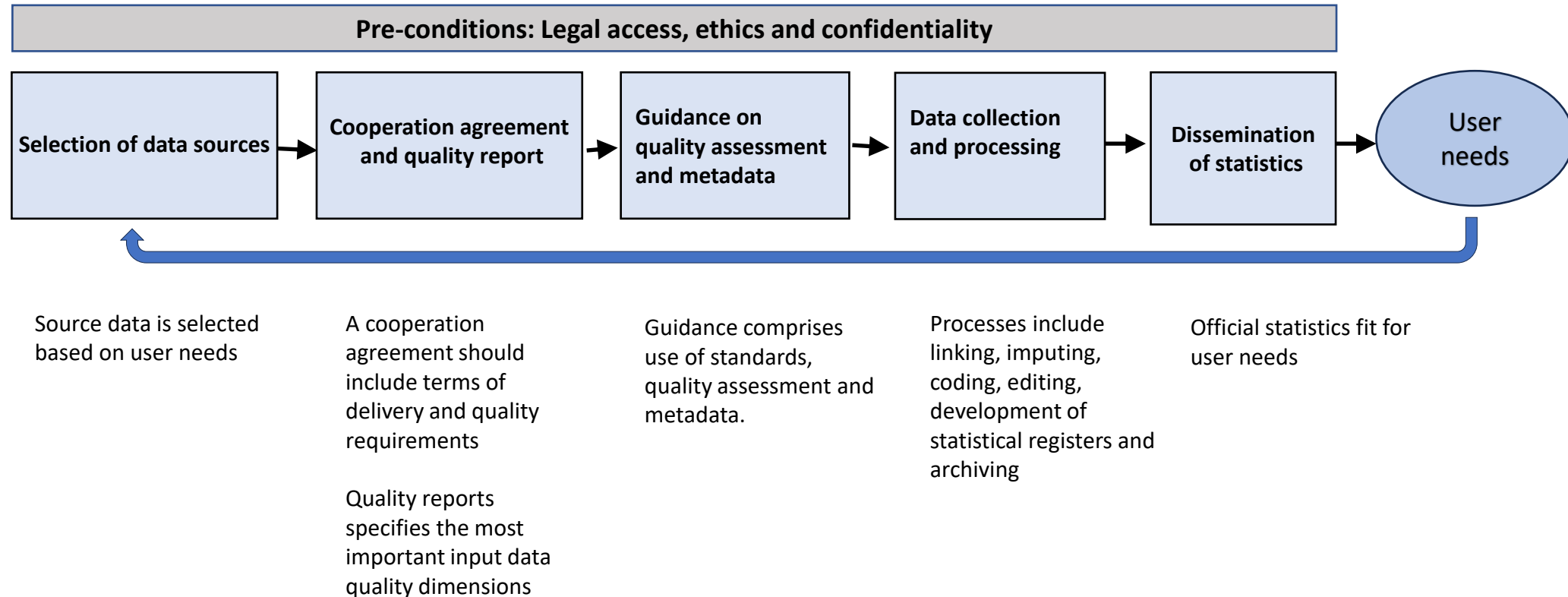
1. Mapping of practices and guidelines to UN NQAF

Source	"Hyperdimension"	Principle / Dimension	Requirements / elements to be assured/ indicators	UN NQAF Principle	Req.	Other Pri	Other Req.
	Framework						
NLD2009	Source	3. Privacy and security	3.1 Legal provision -Basis for existence of data source	2	2.4		2.5
NLD2009	Source	4. Delivery	4.2 Arrangements -Are the terms of delivery documented? - Frequency of delivery	2	2.5	12	12.3
NLD2009	Source		4.3 Punctuality -How punctual can the data source be delivered? - Rate at which exceptions are reported -Rate at which data is stored by data source keeper	2	2.5	12	12.3
NLD2009	Source		4.4 Format -Formats in which the data can be delivered	2	2.5	12	12.3
NLD2009	Source		4.5 Selection -What data can be delivered? - Does this comply with the requirements of NSI?	2	2.5	10	10.3
NLD2009	Source	5. Procedures	5.1 Data collection -Familiarity with the way the data is collected	2	2.5	12	12.5
NLD2009	Source		5.2 Planned changes -Familiarity with planned changes of data source -Ways to communicate changes to NSI	2	2.5	12	12.5
NLD2009	Source		5.3 Feedback -Contact data source keeper in case of trouble? - In which cases and why?	2	2.7	12	12.3
NLD2009	Source		5.4 Fall-back scenario -Dependency risk of NSI delivered according to arrangements made - Emergency measures when data source is not delivered according to arrangements made	2	2.7	12	12.3

1. The most relevant UN NQAF principles on source data

NQAF covers factors contributing to the quality of produced statistics, including practices on handling quality of administrative and other data

Statistical system and Institutional environment	→	Statistical processes	→	Statistical output	→	Users

Quality of statistics

Managing the statistical system

P2 Legal access

P2 Cooperation agreements

P2 Quality reports

P3 Statistical standards

Managing the institutional environment

P5 Ethics

P7 Confidentiality

Managing statistical processes

P10 Selection of sources

P11 Data linking

P12 Assessing the quality of input data

P12 Metadata

P8, P10 – 13 Data collection and processing

Managing statistical outputs

2. Identification of important steps (and requirements) when using administrative and other data for official statistics

Pre-conditions: Legal access, ethics and confidentiality

Selection of data sources → Cooperation agreement and quality report → Guidance on quality assessment and metadata → Data collection and processing → Dissemination of statistics → User needs

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

2. Identification of important steps when using administrative and other data for official statistics – and their link to UN NQAF

Preconditions: Legal access (2.5, 2.6), ethics (5.2) and confidentiality (7)

Selection of data sources (10.3)	Cooperation agreement and quality report (2.5, 2.7, 3.2, 11.5)	Guidance on quality assurance and metadata (2.7, 11.5, 12.3, 12.5)	Data collection and processing (8, 10 – 13)	Dissemination of statistics (14 – 18)	User needs

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

3. Identification of critical requirements

| Source	"Hyperdi mension"	Principle / Dimension	Requirements / elements to be assured/ indicators	UN NQAF Principle	Req.	Other Pri	Other Req.	Identification of critical requirements
NLD2009	Source	3. Privacy and security	3.1 Legal provision -Basis for existence of data source	2	2.4		2.5	Legal basis of source
NLD2009	Source	4. Delivery	4.2 Arrangements -Are the terms of delivery documented? -Frequency of delivery	2	2.5	12	12.3	MoU - Delivery terms
NLD2009	Source		4.3 Punctuality -How punctual can the data source be delivered? - Rate at which exceptions are reported -Rate at which data is stored by data source keeper	2	2.5	12	12.3	MoU - Delivery terms
NLD2009	Source		4.4 Format -Formats in which the data can be delivered	2	2.5	12	12.3	MoU - Data format
NLD2009	Source		4.5 Selection -What data can be delivered? - Does this comply with the requirements of NSI?	2	2.5	10	10.3	Relevance
NLD2009	Source	5. Procedures	5.1 Data collection -Familiarity with the way the data is collected	2	2.5	12	12.5	Metadata about source
NLD2009	Source		5.2 Planned changes -Familiarity with planned changes of data source -Ways to communicate changes to NSI	2	2.5	12	12.5	Metadata about source
NLD2009	Source		5.3 Feedback -Contact data source keeper in case of trouble? - In which cases and why?	2	2.7	12	12.3	MoU - Data cooperation
NLD2009	Source		5.4 Fall-back scenario -Dependency risk of NSI delivered according to arrangements made - Emergency measures when data source is not delivered according to arrangements made	2	2.7	12	12.3	Provider risk and mitigation

4. Identification of critical requirements - 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:

  o **Review of the data holder organization (and its reputation to ensure that it is ethically and legally acceptable to acquire the data)**

  o **Quality assurance of the data holder (beyond assessment of its data)**

  o **Details about the content of Cooperation agreements/MoUs**

  o **Details about the content of quality reports (a checklist on quality dimensions of input data, see next slide)**

  o **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

DESA | **Statistics Division**

5. Review of critical requirements when using administrative and other data sources

- (A) related to selection and/or evaluation of data source

- (B) related to relationship with data provider

- (C) related to use of input data
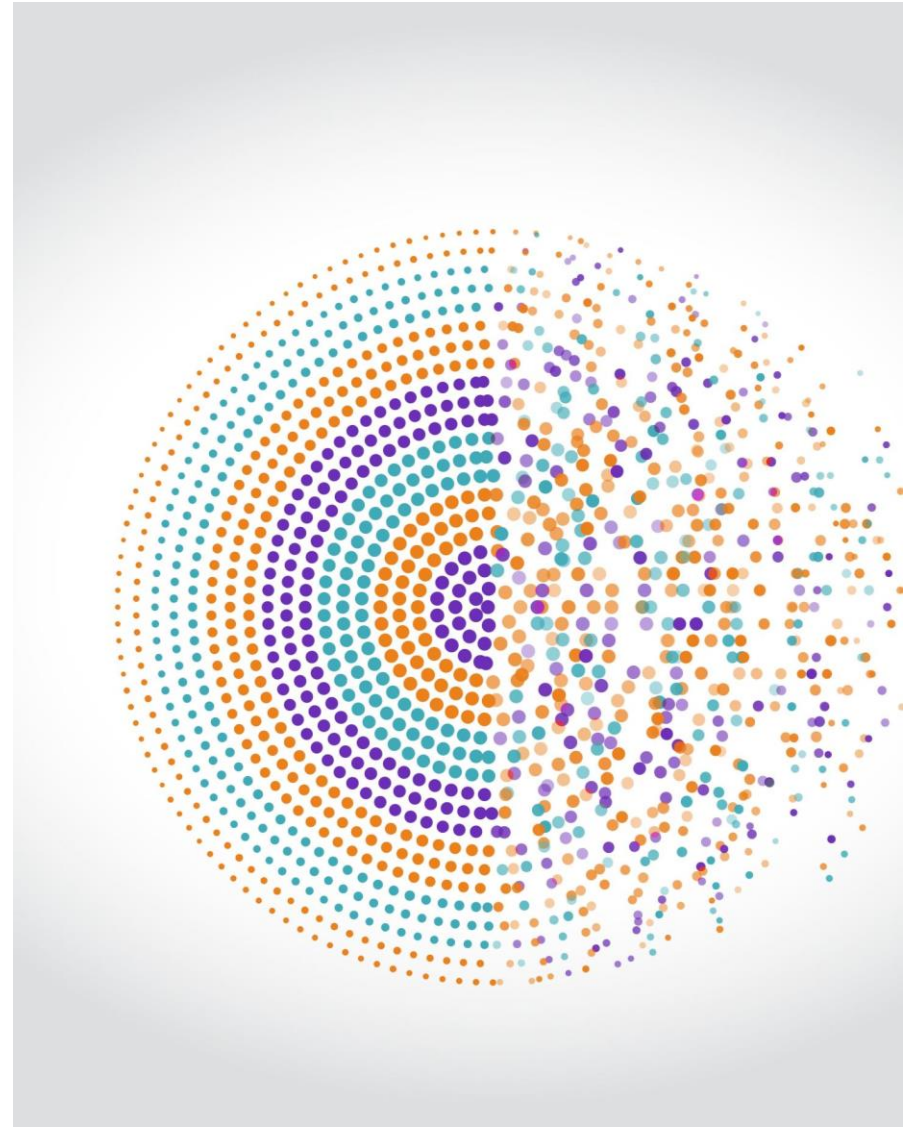
Critical requirements – (A) related to selection and/or evaluation of data source

1. **Data provider information**: There is comprehensive information about the providers of administrative and other data including their name, address, contact person, legal status, residency, management, mandate or purpose

2. **Data provider and source evaluation**: Providers of administrative and other data are assessed in regard to the legal basis of their data source and any risks of the use of their data may provide, and possible mitigation measures

3. **Input data metadata**: There is comprehensive metadata about the administrative and other data regarding their purpose, concepts, definitions and classifications used, units and variables (and possible values), coverage (population), reference area, reference period, timeliness, data structure, data and file format, method of collection, data processing and treatment including for outliers, errors, estimations and imputations), past and future changes over time (stability)

4. **Input data evaluation**: The administrative and other data are thoroughly evaluated during the selection stage according to their potential use and usefulness to produce official statistics, accuracy, completeness, conceptual coherence and comparability, time related dimension (timliness, periodicity and reference period) and accessibility, including cost and confidentiality

Critical requirements – (B) related to relationship with data provider

1. **Data provider and source information**: The statistical agencies have sufficient information about the providers of administrative and other data (metadata about the data source) including how and by whom data is collected, and taking into account the metadata and information collected at the selection and evaluation stage of the administrative and other data.

2. **Cooperation agreement**: There are cooperation agreement between the statistical agencies and the providers of administrative and other data covering access/delivery terms, confidentiality and ongoing cooperation and discussions regarding methodological and quality issues

3. **Provider quality assurance**: The providers of administrative and other data have provided a description of their quality assurance procedures and those procedures are evaluated and found adequate regarding the use of the data for producing official statistics

4. **Input data quality report:** Quality reports for administrative and other data are developed in cooperation with the statistical agencies and the data owner and describe accuracy, completeness, timeliness and punctuality etc.

5. **Data provider and source evaluation**: Providers of administrative and other data are assessed in regard to the legal basis of their data source and any risks of the use of their data may provide, and possible mitigation measures

Critical requirements – (C) related to use of input data

1. **Input data evaluation**: The administrative and other data are evaluated for accuracy, completeness, time related dimension (timeliness, periodicity and reference period), technical checks, and linkability before and at use

2. **Input data metadata**: There is comprehensive metadata about the administrative and other data regarding their concepts, definitions and classifications used, units and variables (and possible values), coverage (population), reference area, reference period, timeliness, data structure, **record count**, data and file format, method of collection, data processing and treatment including outliers, errors, estimations and imputations), past and future changes over time (stability)

3. **Input data processes**: There are well-documented and monitored processes of data pre-treatment including quality controls, error-handling, standardization, data transformation and aggregation, estimation and imputation, data provider consultation and review of possible impact of any conceptual or other changes affecting the administrative and other data sources

(D) Other critical requirements

1. **Legal basis, mandate and responsibility of data provider**

2. **Access to admin and other data (legal and actual)**

3. **Impartiality and objectivity, ethics**

4. **Confidentiality and data security**

Important as well..

1. **Statistical standards**

2. **Quality dimensions of statistical outputs**
   1. Relevance
   2. Accuracy and reliability
   3. Timeliness
   4. Accessibility and clarity
   5. Coherence and comparability

Round table discussion

Do you find those requirements relevant?

/ Do you find anything to be missing?

Thank you.