

Module for Quality Assurance when using Administrative and Other Data Sources to produce Official Statistics

Prepared by the Subgroup on administrative and other data sources
of the United Nations Expert Group on National Quality Assurance Frameworks (EG-NQAF)¹

Draft as of 23 April 2024

Introduction of the module

1. This module distinguishes among three data sources according to their purpose and by the entity responsible for their compilation, statistical data sources such as surveys, administrative data sources, and other data sources. 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. New data sources can often be associated with other data sources; however, they may be considered part of statistical or administrative data sources as well, depending on national circumstances.²
2. The use of **administrative data sources** offers many potential advantages to statistical agencies, such as cost-effectiveness, the reduction of the respondent burden, improved timeliness and improved relevance, accuracy, and reliability due to their ability to obtain highly disaggregated data. The use of **other data sources**, such as available in the private sector can be cost effective, in addition to allowing much more frequent and timely reporting and illustrating phenomena that are difficult or impossible to capture with traditional statistical and administrative data sources. This may lead to improved relevance.
3. This module complements existing generic national quality assurance frameworks (NQAF) and provides more specific and detailed guidance when using administrative and other data sources to produce official statistics. This module cannot be used as a quality assurance framework for administrative and other data sources. It is meant to be used in conjunction with an existing NQAF as, for example, the module misses key aspects of an NQAF, such as related to coordination and the institutional environment.³
4. The module is directed at statistical agencies⁴ that use or want to use administrative and other data sources to produce official statistics. It is based on a review of respective country practices, available guidelines, frameworks and toolkits⁵ and the mapping of their elements to the United Nations National Quality Assurance Framework for Official Statistics (UN NQAF). The

¹ See <https://unstats.un.org/unsd/methodology/dataquality/about/>.

² The United Nations National Quality Assurance Frameworks Manual for Official Statistics (hereinafter referred to as *Manual*) contains a list of other data sources in para. 7.6.

³ Countries without a NQAF can refer to the Roadmap developed by the EG-NQAF which provides detailed guidance on the development and implementation of an NQAF, see <https://unstats.un.org/unsd/methodology/dataquality/roadmap/>.

⁴ See the definition of “statistical agency” in the glossary of Annex 2.

⁵ Please see the repository of country practices, available guidelines, frameworks and toolkits on quality assurance when using administrative and other data sources, available at: <https://unstats.un.org/wiki/pages/viewpage.action?pageId=224264259>.

module consists of two parts, a conceptual approach and a checklist of ten critical requirements. It also includes several annexes with additional information.

5. The conceptual approach in **Part 1** provides an overarching structure of the typical quality considerations when using administrative and other data sources. It consists of preconditions, actions and overarching considerations relating to quality. Those actions and overarching considerations are sometimes referred to as “hyper-dimensions” in various country practices and available guidelines that served as input for this module. **Part 2** provides a checklist of ten critical requirements for assuring quality together with a set of suggested (“best”) practices, which provide more specific and detailed guidance when using administrative and other data sources than what is contained in a generic NQAF. Many of those critical requirements overlap with parts of UN NQAF and other NQAF. The suggested practices identified in this module are ambitious, may not be applicable in all circumstances and may be difficult or not always possible to follow for statistical agencies depending on the specific situation.⁶ However, when applicable, they allow for the identification of areas of improvement in the spirit of the “Plan-Do-Check-Act” cycle for continuous improvement made popular by W. Edwards Deming.

6. **Annex 1** provides a sub-module for input data validation and a list of additional quality indicators. **Annex 2** provides a glossary of terms used in this document. **Annex 3** contains a list of requirements in UN NQAF that are relevant for assuring the quality of official statistics when using administrative and other data sources. **Annex 4** provides a mapping of the ten critical requirements to the relevant UN NQAF requirements listed in Annex 3.

7. The module synthesizes existing country practices, available guidelines, frameworks and toolkits for assuring the quality of statistics when using administrative and other data sources. It aims to be practical and concise. Countries that already have well-established practices for quality assurance when using administrative and other data sources may view this module only as an additional reference point that supports what they are already doing. The module has been developed with reference to UN NQAF, but it is meant to be applicable independently of what NQAF or code of practice for quality assurance is followed by a statistical agency.

8. This module was developed by the Subgroup on administrative and other data sources (hereinafter referred to as Subgroup). The Subgroup was established by the United Nations Expert Group on National Quality Assurance Frameworks (EG-NQAF) to provide practical and concise guidance and best practices for statistical agencies in assuring the quality of official statistics when administrative data sources and other data sources are used for the production of official statistics.⁷ The Expert Group and its Subgroup consist of experts on quality assurance from Member States and international and regional organizations. The Subgroup started its work by reviewing existing country practices and guidelines to identify the quality principles, requirements, elements to be assured and indicators that are critical or of special importance

⁶ For example, no distinction is made on whether a data source is explored for the first time or already acquired and used on a regular basis.

⁷ See terms of reference and membership of the Subgroup, available at: <https://unstats.un.org/wiki/display/EGNQAFSA/EG-NQAF+Subgroup+on+admin+and+other+data+sources>. The mandate of the Subgroup also includes addressing the use of multiple data sources which however is not covered in this module and will be considered separately.

when using administrative and other data sources. These were then mapped to the existing UN NQAF principles and requirements and synthesized into a set of ten critical requirements and a set of suggested practices for each requirement.

Box 1: Definitions of two important terms⁸

National Quality Assurance Framework (NQAF): a coherent and holistic system for statistical quality management. It is a tool for all working in official statistics. Its objective is to achieve quality improvements at the level of the statistical system, including management, coordination and institutional arrangements, processes and statistical outputs in order to meet user needs. It sets a standard of quality and hereby assures trust in official statistics.

United Nations National Quality Assurance Framework (UN NQAF): the generic United Nations (UN) national quality assurance framework, which is contained in Chapter 3 and the Annex of the Manual. The Manual and the recommendations contained therein were adopted by the UN Statistical Commission in March 2019. The UN NQAF consists of principles, requirements, and elements to be assured. The UN NQAF does not aim to replace any of the existing statistical quality assurance frameworks and guidelines for official statistics. Countries and individual producers of official statistics that are already fully engaged in quality assurance and are following one of the existing quality frameworks may view the UN NQAF and the Manual only as an additional reference point that supports what they are already doing, and as a source of information on the application of quality assurance in different situations.

Part 1: Conceptual approach to assure quality when using administrative and other data sources

9. The conceptual approach provides an overarching structure of what typically should be considered to assure quality when using administrative and other data sources. It consists of important preconditions, four actions and two overarching considerations as illustrated in **Figure 1**. Those actions and overarching considerations are referred to as “hyper-dimensions” in some country practices and available guidelines. The four actions can be understood as activities or steps that typically need to take place when using administrative and other data sources. They can follow each other or can be undertaken individually, depending on the specific circumstances. The two overarching considerations reflect aspects that must be considered during the four actions and concern user needs, the quality of input data, and metadata.

10. The conceptual approach is based on the identification and analysis of relevant requirements and practices in a large set of available country practices and available guidelines for quality assurance when using administrative and other data sources and the identification and analysis of relevant requirements and practices in UN NQAF. Accordingly, the actions, overarching considerations and preconditions shown in the conceptual approach in Figure 1 are closely linked to the relevant UN NQAF requirements as indicated by the numbers in brackets.⁹ At the same time, both are linked to the checklist of ten critical requirements in Part 2 as illustrated in Annex 4. The different parts of the conceptual approach are also consistent and

⁸ Based on the *Manual*, available at: <https://unstats.un.org/UNSDWebsite/data-quality/user-manual>.

⁹ Those UN NQAF requirements are listed in Annex 3 for ease of reference.

overlap with the phases and sub-processes of the Generic Statistical Business Process Model (GSBPM), which is applicable to the use of any data source.¹⁰

11. In the following, a brief description of the four actions that are identified in the conceptual approach is provided. The two overarching considerations on user needs, the quality of input data and metadata are included in the description of the four actions, as they take place as part of them and not separately. In addition, there are preconditions that must be considered before or during the process of exploring the use of administrative and other data sources, namely data access and confidentiality.

Action 1: Identification of statistical need and selection of data source

12. The first operational action when using an administrative or other data source is often the selection of the data source itself (see UN NQAF requirement 10.3) based on user needs for relevant statistics (shown as Overarching consideration 1). The selection of the data source also requires a preliminary evaluation of the metadata and input data of the potential data source (Overarching consideration 2). Ethical considerations, the requirement for selecting a data source on an objective basis and an initial cost-benefit analysis are also important when selecting a data source.

Action 2: Cooperation with data providers

13. For a selected data source, a cooperation agreement, or a Memorandum of Understanding with the data provider¹¹ is a good practice to facilitate access and use of observation level and aggregated data (see UN NQAF requirement 2.5 or 2.6). This agreement will contain information on the arrangement of data delivery, such as regarding the technical transfer mechanism, the identification of data sets and data, the timing for delivery and quality requirements. The agreement on a quality report or quality declaration between data provider and statistical agency and referenced or enclosed in the cooperation agreement is a good practice. The national statistical office (NSO) also gives guidance to data providers on statistical standards and cooperates with the data provider on an ongoing basis. The cooperation agreement should be based on a thorough evaluation of the metadata (Overarching consideration 2).

Action 3: Data acquisition and processing

14. The statistical agencies need appropriate procedures for the acquisition and processing of administrative and other data sources (see UN NQAF requirement 12.3). Additional considerations include assuring methodological soundness, improving the potential of administrative and other data sources, assuring appropriate statistical procedures and managing respondent burden through data sharing, data linkage and use of administrative and other data

¹⁰ It is possible to point out GSBPM sub-processes that are especially concerned when assuring the quality of statistics when using administrative and other data sources. For example, the sub-processes 1.1 Identify needs, 1.2 Consult and confirm needs, 1.3 Establish output objectives, and 1.4 Identify concepts correspond to Action 1 of the conceptual approach which is concerned with the selection of the data source based on user needs. However, quality assurance is an overarching process in GSBPM and a detailed level correspondence of quality assurance to GSBPM sub-processes is currently not available.

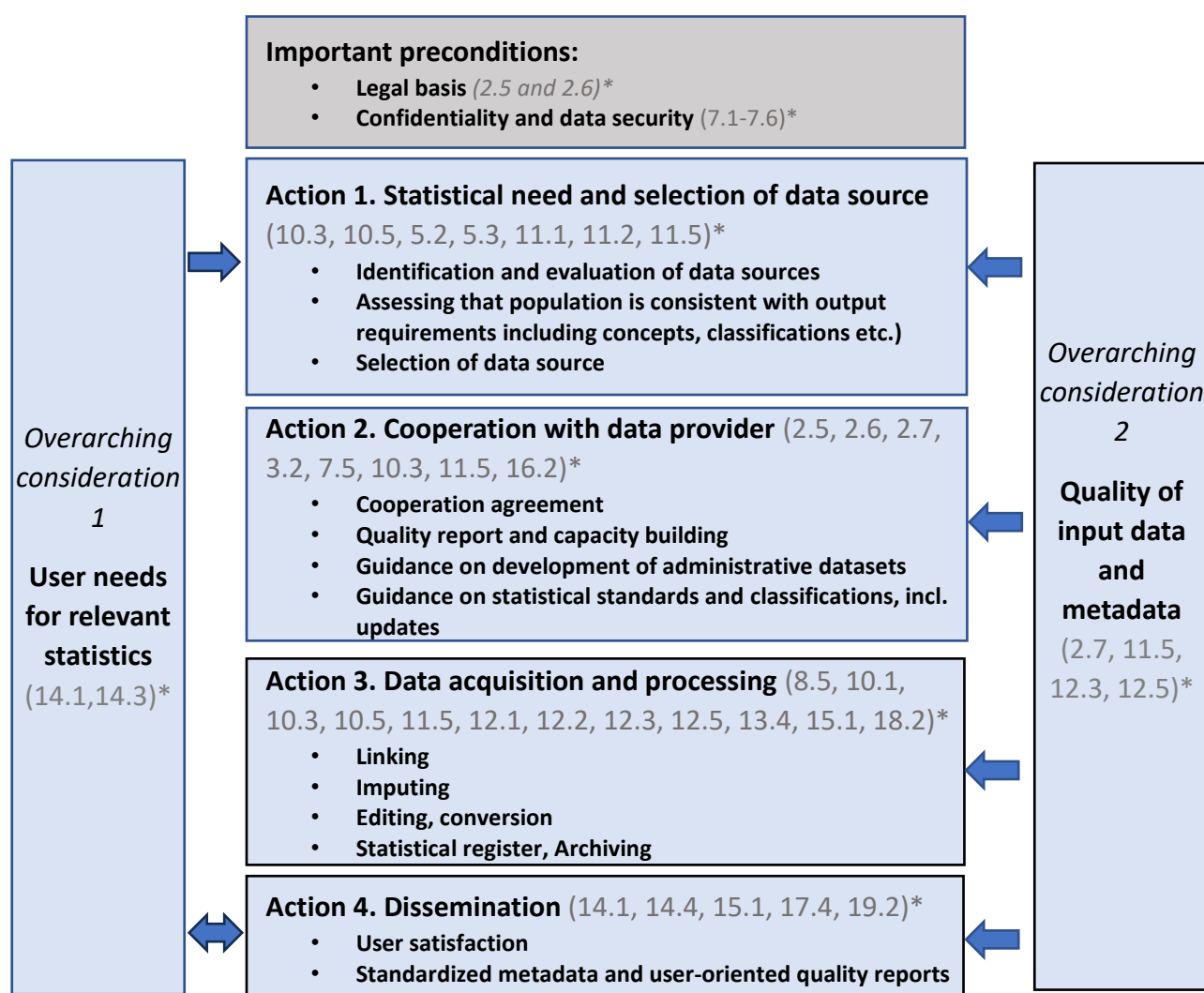
¹¹ See definition of “data provider” in the glossary of Annex 2. Data providers are understood to provide observation level data and aggregated data, as applicable.

sources. The use of input data also requires a thorough assessment of its quality and comprehensive metadata (Overarching consideration 2). Action 3 is closely related to the GSBPM phases of “Collect” and “Process”.

Action 4: Dissemination

15. Users and user needs are identified, and statistics based on new and existing data sources are developed in response to society’s emerging information needs (Overarching consideration 1), and user needs are satisfied. The dissemination of the resulting statistical outputs takes into account the special characteristics of administrative and other data sources.

Figure 1: Assuring the quality of official statistics when using administrative and other input data – a conceptual approach identifying relevant actions and aspects to be considered.



* The numbers in brackets indicate the link to relevant UN NQAF requirements or principles. The relevant UN requirements are listed in Annex 3.

Part 2: Checklist of ten critical requirements

16. Part 2 presents a checklist of ten critical requirements that provide specific and detailed guidance for assuring quality when using administrative and other data sources to produce official statistics. The checklist is complementary to the use of NQAF and is not a substitute for it. The ten requirements are considered critical as without them, the quality of official statistics when using administrative and other data sources may not be sufficiently assured.

17. The ten critical requirements are underpinned by a set of suggested (or “best”) practices that have been identified in a thorough review of a large set of country practices and guidance materials.¹² These suggested practices are often more detailed than the elements to be assured under the UN NQAF requirements. The reason is that UN NQAF is generic, while this checklist aims to provide more specific guidance on the use of administrative and other data sources. Users who have already an established practice of assuring the quality of their official statistics when using administrative and other data sources may find this checklist useful as an additional reference point to validate their existing practices.

18. The checklist of ten critical requirements is closely linked to the actions and overarching considerations of the conceptual approach in Figure 1. The conceptual approach provides an overall structure and understanding for using administrative and other data sources for producing official statistics. The checklist of ten critical requirements provides a concrete tool and practical guidance for this and is complementary to the use of an NQAF. The checklist contains requirements not addressed or not sufficiently reflected in NQAFs when using administrative and other data sources and could be used to update UN NQAF / NQAFs.¹³ The checklist makes no distinction whether a data source is newly explored or already acquired and used on a regular basis.

Box 2: Special case of publicly available data

Many of the ten critical requirements make reference to data providers as the holder or owner of the data of a particular data source that will make their data available for statistical purposes. However, there are also data sources such as social media and data acquisition methods such as web-scraping where existing publicly available data is directly taken by the statistical agency without necessarily engaging in a relationship with the holder or owner of data. Special legal and ethical considerations may apply in such cases, while some of the critical requirements do not apply.

¹² <https://unstats.un.org/wiki/pages/viewpage.action?pageId=224264259>.

¹³ Annex 4 provides a mapping of the ten critical requirements for quality assurance when using administrative and other data sources to the relevant UN NQAF requirements following the conceptual approach. It allows to see what is already covered in UN NQAF and where there are gaps.

TEN CRITICAL REQUIREMENTS

1. **The use of administrative and other data sources must meet a set of preconditions and take user needs into consideration.** This critical requirement summarizes the preconditions of data access and confidentiality, and consideration of user needs that are well reflected in UN NQAF and other commonly used quality assurance frameworks but require special attention when using administrative and other data sources. This critical requirement is reflected in Important preconditions and Overarching consideration 1 of the conceptual approach shown in Figure 1.

Suggested practices

- 1.a. There is legal access to the data.
 - 1.b. There is actual access to the data.
 - 1.c. The data source complies with existing laws and regulations (including the consent of data owners, where applicable) and its data can legally be used for producing official statistics.
 - 1.d. Confidentiality of personal data and business information and data security are assured.
 - 1.e. User needs are taken into account.
2. **New data sources, data providers as well as the use of multiple data sources are proactively explored:** Statistical agencies constantly explore the use of new data sources or the simultaneous use of multiple data sources for producing official statistics. This critical requirement is considered before and when selecting a data source (Action 1 of the conceptual approach).

Suggested practices

- 2.a. There are policies, guidelines, and practical procedures for exploring and testing the potential of new data sources for producing existing and new statistics; this extends to the possible use of multiple data sources through data integration.¹⁴
- 2.b. The statistical agencies work with government bodies, public institutions, private businesses, non-governmental entities and civil-society organizations, academic and research institutions, and other technology innovators to identify and explore the use of new data sources and the integration of multiple data sources.
- 2.c. There is an innovation lab or similar institutional unit, preferably at the NSO that systematically supports the exploration and testing of new data sources and the integration of multiple data sources for producing official statistics across the national statistical office and national statistical system.

¹⁴ Data integration can be achieved, by example, through use of the record linkage methods. Please see Manual, para. 7.18. for some additional information and references.

3. **There is basic information about the data provider and general information about the data source:** There is basic information about the provider of administrative and other data. The statistical agencies have general information about the data source, including how the data is generated. This critical requirement is typically considered when selecting a data source (Action 1 of the conceptual approach).

Suggested practices

- 3.a. There is information about the name and address of the data provider and how to contact the data provider (contact person, organizational unit within the data provider).
 - 3.b. There is information about the legal status, type of organization, residency, purpose or mandate and the management or leadership of the data provider.
 - 3.c. There is general information about the data source, such as the purpose, method and frequency of data collection, data management/storage as well as the population it aims to cover (target population).
 - 3.d. There is a preliminary assessment of the usefulness of the data source for producing official statistics.
 - 3.e. There is general information about the limitations of the data source when used for producing official statistics.
 - 3.f. There is an initial cost-benefit assessment as the use of administrative and other data sources may incur significant costs (see also 7.b).
4. **The data provider and data source are assessed for their risks:** Providers of administrative and other data, as well as the data source itself are assessed concerning any risk the use of the data source may pose, and possible mitigation measures are considered. This critical requirement is considered when selecting a data source or when engaging with the data provider (Action 1 and Action 2 of the conceptual approach).

Suggested practices

- 4.a. There is sufficient information about the data provider, the data source and the data to allow an assessment of the risks the use of this data poses when used for producing official statistics.
- 4.b. The data provider is assessed on whether the use of its data source(s) and data for official statistics pose any risks such as related to lack of continuity and reproducibility of the data compilation, lack of trustworthiness, ethical considerations, risks related to privacy and data security, reputational risks etc.
- 4.c. Measures to mitigate potential risks are identified, including emergency or fallback options when data is not delivered as agreed.

5. **There are cooperation agreements with the data providers and there is ongoing cooperation, as applicable:** There are cooperation agreements between the statistical agencies and the providers of administrative and other data, covering access and delivery terms, confidentiality, data security and ongoing cooperation. This critical requirement is considered when engaging with the data provider (Action 2 of the conceptual approach).

Suggested practices

- 5.a. The statistical agency has cooperation agreements with all its providers of administrative and other data with respect to the use of a particular data source covering both observation level data and aggregates.
 - 5.b. The cooperation agreements with data providers specify the terms of data access including content, coverage, frequency, punctuality and format of data delivery, data security and confidentiality. They include fallback options when data is not delivered as agreed.
 - 5.c. The cooperation agreements specify the ongoing cooperation and communication (feedback and follow-up) arrangements with the data providers regarding the quality of the data and other relevant issues such as changes in the data over time (such as coverage, concepts and definitions, acquisition method etc.) which should be promptly communicated. This practice is also linked to critical requirement 7 that addresses the quality of the input data.
 - 5.d. Statistical agencies are involved in the design, development and processing of administrative data and data from other data sources, as applicable and feasible, in order to make them more suitable for statistical and other purposes, including the data provider's own purposes.
 - 5.e. Data providers are required to respond to any questions and quality issues within a reasonable time while respecting privacy requirements.
 - 5.f. Statistical agencies provide guidance to data providers, including on statistical standards and classifications, as well as any changes of those.
 - 5.g. The cooperation agreement specifies appropriate mechanism for providing feedback on quality issues, any sharing back of data (as permitted by national laws and regulations) and other benefits or services the statistical agency can provide to the data provider.
 - 5.h. There is a central body at the NSO or within the NSS that supports the establishment of cooperation agreements with data providers.
6. **The data provider assures the quality of its data and produces a quality report (or quality declaration) in cooperation with the statistical agency, as applicable:** The data provider assures the quality of the source data for its intended purposes and provides a quality report (or similar) to the statistical agency. This critical requirement is considered when engaging with the data provider (Action 2 of the conceptual approach) and is part of Overarching consideration 2 of the conceptual approach.

Suggested practices

- 6.a. The providers of administrative and other data have provided a description of their quality assurance procedures when compiling and processing the data (including any transformations, data editing, estimation, dealing with missing units, values and outliers, etc.) as applicable and feasible.
 - 6.b. The quality assurance procedures of the data providers are evaluated and the resulting data is found to be, in general, adequate regarding its use for producing official statistics.
 - 6.c. A quality report, quality declaration or note describing the accuracy, completeness, timeliness, punctuality and other relevant characteristics of the data is developed in cooperation between the statistical agency and the data provider.
 - 6.d. The results of quality audits and other forms of quality assessments conducted by or at the data provider are shared with the statistical agency if available and as appropriate and feasible.
7. **The quality of the input data is systematically evaluated by the statistical agency:**
The quality of the input data is systematically and regularly assessed and evaluated at all relevant stages of the statistical production process. This critical requirement is considered when selecting a data source (Action 1 of the conceptual approach), when using the data (Action 3 of the conceptual approach) and as part of Overarching consideration 2 of the conceptual approach.

Suggested practices

- 7.a. The administrative and other data are systematically evaluated during the selection stage in cooperation with the data provider, according to a set of predefined objective criteria for their potential use and usefulness to produce official statistics, including according to accuracy, completeness, coverage of different groups and possible bias due to the under-, or over-representation of specific groups, conceptual coherence and comparability, time-related dimension (timeliness, periodicity and reference period) and accessibility, including cost and confidentiality.
- 7.b. The selection of a data source follows a cost-benefit analysis considering human resources, infrastructure and other costs and sustainability as the use of administrative and other data sources may incur significant costs (for example, in the case of very large raw datasets).
- 7.c. The administrative and other data are systematically and regularly evaluated before and during use for accuracy, completeness, coverage of different groups and possible bias due to the under-, or over-representation of specific groups, conceptual coherence and comparability, time-related dimension (timeliness, periodicity and reference period), technical checks/accessibility, and integrability/linkability and the results are reflected in periodic quality reports (see Annex 1, Table 1: Sub-module for input data validation).

8. **There is comprehensive metadata about the input data:** There is comprehensive and standardized metadata about the administrative and other data regarding concepts, definitions and classifications used, data structure, units and variables (and possible values), coverage (population), reference area, reference period, timeliness, method of collection, data processing and treatment including for outliers, errors, estimations and imputations, record count, data and file format, past and future changes over time (stability), etc. This critical requirement is considered when selecting a data source (Action 1 of the conceptual approach), when using and disseminating the data (Action 3 and Action 4 of the conceptual approach) and as part of Overarching consideration 2 of the conceptual approach.

Suggested practices

- 8.a. There is comprehensive information about the concepts, definitions and classifications used.
 - 8.b. There is comprehensive information about the dataset, including the data structure, the units and variables (and possible values and code sets), coverage (population), reference area, reference period and timeliness.
 - 8.c. There is detailed information about the basis for and method of data collection, data processing and data editing including for outliers, errors, estimations, imputations and use of algorithms, software and AI at the data provider, and any rules and regulations related to the generation of the data, as applicable.
 - 8.d. There is information about the record count, data and file format.
 - 8.e. The metadata follows a standard format.¹⁵
 - 8.f. There is comprehensive information about relevant past and planned changes to the data compilation over time (stability).
 - 8.g. Possible limitations of the data source for producing official statistics are identified, including those related to population coverage, completeness of information, possible bias, and alignment of administrative units with statistical units of interest (e.g., tax unit vs. establishment or enterprise).
9. **Processing of input data at the statistical agency follows standards, guidelines, and best practices:** Data processing and data editing follows standards, guidelines and best practices, and is documented and monitored. There is an ongoing collaboration with the data provider. This critical requirement is considered when using the data (Action 3 of the conceptual approach) and part of Overarching consideration 2 of the conceptual approach.

¹⁵ A standard format for input data metadata can be developed based on existing related standards such as the European Statistical System's standard on reference metadata and quality reporting for statistical processes and outputs; see European Statistical System (ESS) Handbook for Quality and Metadata Reports — re-edition 2021, available at: <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-21-021>.

Suggested practices

- 9.a. There are guidelines for quality assurance and management at the statistical agency when using administrative and other data sources that should be standardized as much as possible and also reflect the specific circumstances.
 - 9.b. The Generic Statistical Business Process Model is used to structure and document the statistical process.
 - 9.c. The processes of data pre-treatment and treatment including quality controls, error-handling, standardization, data transformation and aggregation, estimation and imputation at the statistical agency are well-documented, tested and monitored, and follow best practices.
 - 9.d. The impact of any conceptual or other changes affecting the administrative and other data sources is systematically evaluated by the statistical agency.
 - 9.e. Data providers are consulted by the statistical agency in case of any questions and quality issues.
 - 9.f. The quality of statistical outputs including statistical registers based on administrative and other data sources is systematically assessed by the statistical agency.
 - 9.g. The use of administrative and other data sources is promoted and practiced at the statistical agency, including through data sharing with other organizational units and other statistical agencies, and through data linkage, using advanced tools, technologies and methods, including SDMX.
10. **The dissemination of statistical outputs meets the needs of users.** The special characteristics of administrative and other data sources are considered when disseminating statistical outputs. This critical requirement is considered in the dissemination of statistical outputs (Action 4 of the conceptual approach) and as part of Overarching considerations 1 and 2 of the conceptual approach.

Suggested practices

- 10.a. The use of administrative and other data sources is based on society's emerging information needs, including consultation with users; user satisfaction is regularly measured and followed up on.
- 10.b. Users receive standardized metadata and quality reports or statements that inform about the use of administrative and other data sources and the relevant quality aspects as well as the limitations of the statistical output, including accuracy, reliability, coherence and comparability.
- 10.c. Quality indicators are suited to inform about the quality of statistics based on administrative and other data sources.
- 10.d. Access to microdata from administrative and other data sources is subject to the statistical agencies' rules and protocols on statistical confidentiality unless different arrangements have been agreed with competent authorities and the data provider and are made available to the public as applicable.

Annex 1: Sub-module for input data validation and list of additional indicators to assess the ten critical requirements

Table 1 provides a sub-module for input data validation once the input data is received by the statistical agency from the data provider.¹⁶ The suggested quality indicators may be part of a quality report that the statistical agency produces every time it receives data. Most of the checks can be automated and are part of what is typically referred to as structural validation of a dataset.¹⁷ This input data validation report can be shared with the data provider. Depending on the specific situation and cooperation agreement, there could be further steps towards an integration of quality assurance at the data provider and statistical agency that could also include cooperation on how to address specific quality issues. The indicators on accuracy, completeness and representativeness and indicator 23 on the comparability of units are candidates for inclusion in quality reports for users.

Table 1: Sub-module for input data validation: Possible (numerical) indicators*

Quality dimensions / Object	Indicators	Metrics
Technical checks/accessibility		
- Data set	1. Readability/accessibility: Data set of the source is accessible and machine-readable	Yes/No
	2. File structure and metadata compliance: Data set contains all expected fields	Yes/No/Unknown
	3. Number of record count: Number of records received divided by expected (e.g., last submission's) number of records	Percentage
- Variables	4. Number of variable counts: Number of records with valid value for a variable divided by number of records	Percentage
Accuracy (numerical measures)		
- Units (objects) of observation	5. Legitimacy/authenticity: Units without allowed (wrong or invalid) identification key(s), if applicable, with percentage calculated based on the number of observations in the data set	Percentage
	6. Inconsistency in data set: Units with non-logical relationships with other units, if applicable, with percentage calculated based on the number of observations in the data set (correction is mandatory before proceeding)	Percentage
	7. Implausibility in data set: Units with implausible or suspicious relationships with other units, if applicable, with percentage calculated based on the number of observations in the data set (correction is optional before proceeding)	Percentage

¹⁶ The list of proposed indicators is already widely used for administrative data, and more experience and feedback are required to better assess their applicability for other data and to develop additional indicators, if needed.

¹⁷ For further information on the implementation of automated validations please refer to the work of the SDMX Secretariat Task force for the Validation and Transformation Language (VTL) which developed a standard language for defining validation and transformation rules, see https://sdmx.org/?page_id=5096.

Quality dimensions / Object	Indicators	Metrics
- Variables	8. Measurement error: Values for which a measurement error is marked by the data provider, with the percentage calculated based on the number of observations in the data set	Percentage
	9. Inconsistent values: Values with non-logical relationship with other information for the unit of observation or outside of possible range (errors), with the percentage calculated based on the number of observations in the data set (correction is mandatory before proceeding)	Percentage
	10. Implausible values: Values with implausible or suspicious relationships with other information of the unit of observation or outside of the expected range (outliers), with percentage calculated based on the number of observations in the data set (correction is optional before proceeding)	Percentage
Completeness / Representativeness		
- Units (objects) of observation	11. Missing units: Units not included, with the percentage calculated based on the number of observations in target population	Percentage
	12. Units outside of the target population: Units that do not belong to the target population with the percentage calculated based on the number of observations in data set	Percentage
	13. Redundancy/duplicates: Units that are duplicate when no duplicate units are expected, with the percentage calculated based on the number of observations in the data set	Percentage
	14. Selectivity/representativity: Deviations (over and under coverage) in relevant characteristics between the data set and the target population, e.g., the percentage of characteristics in the data set minus percentage in the target population	Percentage
- Variables	15. Missing values, with the percentage calculated based on the number of observations in the data set	Percentage
	16. Imputed values: Values for which an imputed value is marked by the data provider, with the percentage calculated based on the number of observations in the data set	Percentage
Time related dimension		
- Data set	17. Timeliness: Date of receipt – Date of end of reference period	Days
	18. Average delay in registration: Date of registration of administrative data entries – Date of end of reference period	Days
	19. Punctuality: Date of receipt – Date of agreed delivery	Days
- Units (objects) of observation	20. Dynamics: New units in data set, Previous units not existing in the data set anymore, with the percentages calculated based on the number of observations in the data set	Percentages
- Variables	21. Stability: Change in the values of variables of existing units over time	Percentages
Linkability / integrability		
- Data set	22. Linkability: Units that can be clearly linked to units in the register or other data sets (e.g., via linking variable or unique identifier), with the percentage calculated based on number of observations in data set	Percentage

Quality dimensions / Object	Indicators	Metrics
- Units (objects) of observation	23. Comparability of units in source: Units with the desired concept definition, with the percentage calculated based on number of observations in data set	Percentage
- Variables	24. Comparability of values: Differences in the value of variables of linked units	Percentages

* Adapted from Daas P, Ossen S., BLUE-ETS (2011): Deliverable 4.2: Report on methods preferred for the quality indicators of administrative data sources. Available at: http://www.pietdaas.nl/beta/pubs/pubs/BLUE-ETS_WP4_Del2.pdf.

Most of the suggested (“best”) practices in the checklist of ten critical requirements are formulated to allow a yes/no/partially response and can therefore serve as quality indicators, or quality indicator checklist. **Table 2** provides an illustrative list of examples of additional indicators mostly taken from a list of quality indicators developed for the Generic Statistical Business Process Model (GSBPM). Many indicators in table 2 overlap with the suggested (“best”) practices in the checklist of ten critical requirements. Further work is needed to develop a consolidated indicator checklist for assessing the status of implementation of the ten critical requirements.

Table 2: Illustrative list of additional indicators for critical requirements*

Critical requirement	Indicators*	Possible metrics
CR 1	1. To what extent have stakeholders confirmed the detailed statistical needs (what, when, how and why) as documented by the NSO?	Yes/No/Partially
	2. To what extent does the data source satisfy information demand?	Yes/No/Partially
CR 4	3. Expected length of comparable time series.	Number of periods
CR 5	4. Existence of an advance notification plan about the forthcoming changes to the data source. Is a contingency plan for changes to the data or data source in place?	Yes/No
	5. Percentage of data transmitted according to the agreements with administrative data owners (e.g., format, time schedule) with the percentage calculated based on number of observations in data set.	Percentage
CR 7	6. Extent to which concepts, definitions and classifications associated to (key) variables and populations follow international or national standards.	Scale
	7. Degree of conceptual compliance of actual unit (object) of observation with desired unit of observation and the impact of any difference on the accuracy of data**	Qualitative
	8. Degree of conceptual compliance of variables with desired variables and the impact of any difference on the accuracy of data**	Qualitative
CR 8	9. When has the data collection technique last been revised/improved?	Years
CR 9	10. Extent to which administrative data integration techniques are understood and specified, both for direct and indirect use of administrative data sets.	Yes/No/Partially
	11. To what extent is the business process using standard or well-known methods for subsequent phases (e.g., coding,	Yes/No/Partially

	editing and imputation, data integration, weighting, estimation, revision), in a transparent way?	
	12. Percentage of identified and documented GSBPM processes (with sub-processes) with their flows	Percentage
	13. Specifications for production systems and workflow take into consideration the type of data being processed (respondent data or administrative data or a combination).	Yes/No/Partially
	14. Have process components for data linkage been tested and fine-tuned?	Yes/No/Partially
	15. Degree of coherence with other sources, with provisional data, with quick estimates, and with previous results of the same process.	Scale
	16. Comparability of results over time and with results from other data sources**	Qualitative
CR 10	17. To what extent have legal constraints been considered regarding statistical outputs, for example but not limited to, ensuring confidentiality of data and preventing the disclosure of sensitive information?	Yes/No/Partially
	18. Have the confidentiality rules and micro data access procedures been designed?	Yes/No/Partially
	19. To what extent is the data actually protected? What is the residual risk of disclosure?	Qualitative

* Selected from United Nations Economic Commission for Europe, Quality Indicators for the Generic Statistical Business Process Model (GSBPM) - For Statistics derived from Surveys and Administrative Data Sources, Version 2.0, October 2017.

** Added indicators.

Annex 2: Glossary of working definition of relevant terms

This glossary provides definitions of important terms used in this module for the quality assurance when using administrative and other data sources. The definitions presented below are used throughout this module, but countries may have their own definitions. Some of the definitions describing quality aspects must be applied to the specific situation, e.g., timeliness is not the same for producers of official statistics that use input data and for users of statistical outputs. The below definitions are taken from the United Nations National Quality Assurance Frameworks Manual for Official Statistics (hereafter referred to as Manual) (see para. 1.14) unless indicated otherwise.

Terms that describe quality aspects

- **Accuracy:** the closeness of estimates to the exact or true values that the statistics were intended to measure.
- **Accessibility:** the ease and conditions with which statistical information can be obtained.
- **Clarity:** the availability of appropriate documentation relating to the statistics and the additional assistance that producers make available to users.
- **Coherence and consistency:** the ability to reliably combine statistics and data sets in different ways and for various uses. Consistency is often used as a synonym for coherence.¹⁸
- **Comparability:** the extent to which differences in statistics from different geographical areas, non-geographical domains, or over time, can be attributed to differences between the true values of the statistics.
- **Completeness and coverage:** Completeness refers to the extent to which all statistics that are needed are available. The measurement of the availability of the necessary statistics normally refers to data sets [set of observations] and compares the required data set to the available one. Coverage is the definition of the scope of the data compiled. This metadata element is used to describe the dimensions delimiting the statistics produced, e.g., geographical, products, economic and other sectors, industry, occupation, transactions, demographic groups. etc., as well as relevant exceptions and exclusions. It can also specify the period of time for which data are provided (see SDMX Glossary – Version 2.1 – December 2020).
- **Confidentiality and privacy:** property of data indicating whether they are subject to dissemination restrictions. Data are protected by confidentiality in cases where unauthorized disclosure could be prejudicial or harmful to the interest of the source or other relevant parties. For instance, data allowing the identification of a physical or legal person, either directly or indirectly, may be characterized as confidential according to the relevant national or international legislation (see for further details SDMX Glossary – Version 2.1 – December 2020). Confidential data means data that allow individual statistical units to be identified, either directly or indirectly, thereby disclosing individual information. To determine whether a statistical unit is identifiable, account shall be taken of all relevant means that might reasonably be used by a third party to identify the statistical unit (see European Statistical System handbook for quality and metadata reports, 2020 edition). Privacy is the state of being alone and not watched or disturbed by other people (Oxford Dictionary).

¹⁸ However, some statistical agencies make a distinction between both terms, with coherence referring to the use of standards and consistency referring to the internal logic of a data set.

- **Integrability / Linkability:** the ease by which the data in the source can be integrated into the statistical production system. For statistical units (objects) it is the comparability and ease of linking of the units in the source to those commonly used by the statistical agencies. For variables it is the closeness of the values in the source to the facts of similar variables (see Daas P, Ossen S., BLUE-ETS (2011)).
- **Relevance:** the extent to which the statistics satisfy the needs of the users.
- **Reliability:** the closeness of the initially estimated value(s) to the subsequent estimated value(s) if preliminary figures are disseminated.
- **Timeliness:** the length of time between the end of a reference period (or date) and the dissemination of the statistics.
- **Punctuality:** the time lag between the release date and the target date by which the data or statistics should have been delivered.

General terms

- **Data providers and statistics producers:** the Manual distinguishes between data providers, who provide an input to the statistics production process (such as respondents and holders or owners of statistical, administrative and other forms of data), and statistics producers, who produce a statistical output. Depending on the specific context, when using the term data provider the Manual refers only to holders or owners of data.
- **Data sources:** the Manual distinguishes among three data sources according to their purpose and by the entity responsible for their compilation: statistical data sources such as surveys; administrative data sources; and other data sources. 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. New data sources can often be associated with other data sources; however, they may be considered part of statistical or administrative data sources as well, depending on national circumstances. - The Manual, para. 7.6 contains a list of other data sources.
- **Data set, data structure, observations, units and variables:** a data set refers to an organized collection of observations that share the same data structure (dimensions, attributes and measures). Observations may also be referred to as “records” or “data records”. Observations contain information about the unit or object of observation which could be a person, a household, a business, a location, transaction, etc. A variable is a characteristic of a unit being observed that may assume more than one set of values (internal working definition partially based on SDMX Glossary – Version 2.1 – December 2020).
- **Input data and source data:** refers to the data that is used in the statistical production process. In the context of this module, input data and source data mean the same (internal working definition).
- **Metadata:** data that define and describe other data. Structural metadata and reference metadata can be distinguished from each other. Structural metadata define and accompany the data and consist of identifiers and descriptors that are essential for discovering, organizing, retrieving and processing a statistical data set (e.g., titles, subtitles, short descriptions, dimension names, variable names, etc.) Reference metadata are of a more general nature and describe statistical concepts and methodologies used for the collection and generation of data and provide information on data quality, thereby assisting users with the interpretation of the data. Contrary to structural metadata, reference metadata can be decoupled from the data (i.e., they can be generated, collected, or disseminated separately from the statistics to which they refer).

- **National quality assurance framework (NQAF):** a coherent and holistic system for statistical quality management that assures trust in and the quality of official statistics. It is a tool for all working in official statistics.
- **National Statistical Office (NSO):** the leading statistical agency within a national statistical system. National statistical office and national statistical institute mean the same thing. In general, the NSO has a coordination role within the national statistical system, and is responsible for the development, production, and dissemination of official statistics across multiple statistical domains.
- **National statistical system (NSS):** the ensemble of statistical organizations and units (statistical agencies) within a country that develop, produce and disseminate official statistics on behalf of the national Government (and other levels of government). It is the responsibility of each country to define the scope of its NSS (see also statistical agencies, data providers and statistics producers).
- **Official statistics:** statistics that 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 NSS in compliance with the Fundamental Principles of Official Statistics and accepted quality frameworks such as the UN-NQAF, as well as other internationally agreed statistical standards and recommendations. In many countries, official statistics are defined and described in the statistical programmes.
- **Quality report:** a typical way of recording the results of a quality assessment. One can distinguish a producer-oriented report comprising metadata to record quality problems and improvements vis-a-vis a user-oriented report comprising metadata that are intended for users of the statistical outputs, enabling them to assess whether the outputs are appropriate for the purposes they have in mind (see European Statistical System handbook for quality and metadata reports, 2020 edition).
- **Statistical agencies:** members of the NSS, encompassing the NSO and other producers of official statistics. Statistical agencies other than the NSO normally have other main purposes and tasks than the production of official statistics and only a section or a small group of people within the institution produces statistics. The quality requirements for processes and output are the same for all official statistics. However, for a ministry or administrative body where only a part of that body produces statistics, the requirements linked to the institutional environment apply only to the entity producing official statistics. For example, while the ministry or administrative body is typically not independent, the unit within the ministries/administrative bodies that is responsible for producing statistics should decide on how to produce and when to disseminate its statistics independently.

Annex 3: Relevant UN NQAF requirements

Annex 3 provides the complete text of the requirements of UN NQAF (contained in the Annex of the Manual) that are relevant or contain suggested (“best”) practices or elements to be assured that are relevant for assuring the quality of administrative and other data sources. The most relevant requirements and elements to be assured are highlighted. The requirements are listed as they are and in their sequence in UN NQAF without any changes. Annex 3 is provided solely for ease of reference, also having in mind users in countries or statistical agencies that do not use UN NQAF.

Requirement 2.5: The national statistical office and, if appropriate, other statistical agencies have the legal authority or some other formal provision to obtain administrative data and adequate access to those data from other government agencies for statistical purposes.

- The statistical law provides appropriate provisions to guarantee the NSO and, if appropriate, other statistical agencies the right to obtain or access administrative data in a timely manner.
- Where statistical agencies do not have a legal right to obtain administrative data, memorandums of understanding are in place that provide such access.
- Statistical agencies’ access to administrative data are free of charge.
- Agreements with owners of administrative data are in place to operationalize data access which describe technical conditions for access and possibilities for linking the data with data from other administrative data sources.
- Statistical agencies are involved in the design and development of administrative data sets in order to make them suitable for statistical purposes; this involvement extends to the possible discontinuation of such data sets.

Requirement 2.6: The national statistical office and, if appropriate, other statistical agencies have the legal authority or some other formal provision and related agreements to access and use 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.

- The statistical law provides appropriate provisions to guarantee the NSO and, if appropriate, other statistical agencies the right to obtain or access, in a timely manner, data held by private corporations or other non-governmental organizations for statistical purposes (e.g., all corporations that provide services to individuals and legal entities residing in the country).
- The statistical law foresees adequate sanctions to ensure access to privately held data where appropriate (such as fines for not granting such access).
- Where statistical agencies do not have a legal right to obtain access to data maintained by corporations or other non-governmental organizations, memorandums of understanding are in place that provide such access.
- Statistical agencies consider the relevance and the scope of data requested.
- The access and use of privately held data follow procedures agreed between the statistical agencies and the owners or holders of the data.

Requirement 2.7: The national statistical office cooperates with and provides support and guidance to data providers.

- The NSO regularly consults with data providers and maintains cooperation with the providers of administrative data and with corporations, businesses and other organizations that hold data to strengthen the statistical value and usage of these sources.
- Quality reports for administrative data are developed in cooperation with the NSO and the data owner and describe accuracy, completeness, timeliness and punctuality, among other things.⁹⁰
- Holders of administrative data, businesses and other organizations receive feedback on the quality of the data provided, allowing for further improvements.

- Partnership agreements with data providers are in place.

Requirement 3.2: The national statistical office provides support and guidance to all data providers and producers of official statistics in the implementation of statistical standards.

- The NSO monitors the extent to which statistical standards are used by data providers and producers of official statistics.
- Periodic reports are prepared with regard to compliance with international, regional and national statistical standards.
- Statistical standards are communicated and made available to all data providers and producers of official statistics.
- Plans and schedules for the development and application of new standards are communicated in advance.
- The NSO assists other statistics producers and data providers in the implementation of international, regional and national statistical standards as appropriate.

Requirement 5.2: The statistical agencies implement a declaration or code of conduct or ethics which governs statistical practices, and compliance with it is followed up.

- There are ethical guidelines or a code of conduct for assuring impartiality and objectivity.
- The guidelines are available to the public.
- The implementation of the guidelines is followed up.

Requirement 5.3: Data sources and methodologies are chosen on an objective basis.

- Sources, concepts, methods and processes for the development, production and dissemination of data are chosen on the basis of statistical considerations, national and international principles and best practices.

Requirement 7.1: Statistical confidentiality is guaranteed by law.

- There is a law or some other clear formal provision in force that mandates the proper management of information received from respondents and data providers to ensure statistical confidentiality and data security.

Requirement 7.2: Appropriate standards, guidelines, practices and procedures are in place to ensure statistical confidentiality.

- Guidelines and instructions on the protection of statistical confidentiality throughout the statistical business process are provided to all staff of the statistical agencies.
- There are regular and continuous training programmes for all staff on the concept of statistical confidentiality and best practices to ensure the privacy of the information provided.
- The organizational structure and arrangements for the development and implementation of practices for ensuring statistical confidentiality is adequate to cope with needs.
- Staff sign confidentiality agreements upon their appointment, which are also valid after staff leave the agency.

Requirement 7.3: Strict protocols to safeguard data confidentiality apply to users with access to microdata for research or statistical purposes.

- Clear conditions for granting access by researchers to confidential data for scientific purposes are set in the statistical law or other formal provision.
- Confidentiality rules, disclosure control and microdata access procedures apply throughout the statistical business process.
- The statistical agencies monitor the use of microdata sets to identify any circumstances in which data confidentiality may be breached (e.g., through file matching), and take immediate corrective action to address such a situation.

Requirement 7.4: Penalties are prescribed for any wilful breaches of statistical confidentiality.

- Legal or other provisions are in place that allow administrative, penal and disciplinary sanctions for the violation of statistical confidentiality.
- Information on the provisions that allow sanctions for the violation of statistical confidentiality is shared with all staff and is available to the public.

Requirement 7.5: The security and integrity of data and their transmission is guaranteed by appropriate policies and practices.

- An IT security policy is in place and is known to the staff.
- Following the IT policy, appropriate physical security measures and processes are in place to ensure data and database security, in accordance with best practices and international standards.
- Regular security audits of the data security system are carried out.
- All access to data repositories and transmission channels is monitored.
- While data are being transferred, risk of a breach is assessed and appropriate procedures are applied to eliminate or minimize this risk.

Requirement 7.6: The risk that individual respondents may be identified is assessed and managed.

- There should be a balance between the acceptable level of risk of identification of individual respondents and the usability of the data.
- Appropriate processes are in place to assess the risk of disclosure of sensitive information and the risk that individual respondents can be identified from the public release of statistics or of microdata, and procedures are applied in line with the data dissemination policy to minimize this risk.
- All procedures taken to adequately reduce the risk of identification are properly documented and made available as part of the metadata related to the statistical data set.
- Users are made aware that procedures to reduce the risk of identification have been implemented and that such procedures could lead to a loss of information.

Requirement 8.5: Guidelines for implementing quality management are defined and made available to the public.

- Guidelines for implementing quality management are produced and issued which:
 - Describe the quality principles and framework followed
 - Describe the entire statistical process and identify relevant documentation for each stage of production
 - Describe the methods for monitoring the quality at each stage of the statistical production process
 - Identify the indicators (quality measures) for evaluating the quality of the main stages of production, including indicators for source data
- The guidelines, methodological manuals and handbooks on recommended practices for quality assurance are made available to the public.
- Mechanisms are in place to assure the quality of data collection (including the use of administrative data and other sources) and data editing.

Requirement 10.1: The methodologies applied by the statistical agencies are consistent with international standards, guidelines and good practices and are regularly reviewed and revised as needed.

- Organizational structures for the development and application of sound statistical methods are commensurate to needs.
- Review and reporting processes are in place that allow the management of the statistical agency to be assured that sound methodological approaches have been adopted and applied throughout the production process.
- The methodologies of surveys and the use of administrative data and other sources of data are evaluated periodically.
- Sampling design is based on sound methodology.
- Proper follow-up procedures are planned and implemented in cases of nonresponse.
- Statistical editing procedures and imputation methods are based on sound methodology.
- When statistical modelling is used in the statistical production process (e.g., for seasonal adjustment), the validity of model assumptions is carefully considered and the impact on final estimates is evaluated.
- Statistical agencies review the methods used by external partners for the compilation of data and the production of statistics.

Requirement 10.3: The statistical agencies choose data sources taking into account accuracy and reliability, timeliness, cost, the burden on respondents and other necessary considerations.

- The use of alternative sources of data, including existing surveys and census, administrative data, big data or other sources of data, is constantly evaluated.
- Quality has to be assessed when using administrative data or other data sources. Ideally, when using administrative data, it should be assured that:
 - The population is consistent with the statistical output requirements
 - The classifications are appropriate
 - The underlying concepts are appropriate
 - The records are complete and up to date
 - The geographical coverage is complete and the measurement units are appropriately defined/identified
- When using other data sources (such as big data), the specific methodological challenges such as those linked to the statistical population and the veracity and volatility of such data have to be considered.

Requirement 10.5: The statistical agencies cooperate with the scientific community to improve methods and promote innovation in the development, production and dissemination of statistics.

- Collaboration with the scientific community is in place, for example through conferences, workshops, task forces and training/courses, to discuss relevant methodological and technological developments (e.g., with regard to exploiting new data sources).
- There are agreements in place with academic institutions on cooperation and the exchange of qualified personnel.
- Staff collaborate on methodological issues with colleagues at the international level.
- Regular participation and presentations at relevant national and international conferences is encouraged for the exchange of knowledge and experiences.
- National and international conferences, seminars, workshops or similar events with the participation of the scientific community are organized by the statistical agencies.

Requirement 11.1: The costs of producing all individual statistics are measured and analysed, and mechanisms are in place to assure the cost-effectiveness of statistical activities or processes.

- There is a system for registering cost and time used for all statistical products, and estimating time used on the main processes should be possible.

- The costs of producing the statistics are well documented at each stage of the production process and are regularly reviewed and analyzed across statistical products to assess the effectiveness of their production.
- Cost-benefit analyses are carried out to determine the appropriate trade-offs in terms of data quality.
- The cost-effectiveness of every statistical survey is assessed.
- The need for each survey variable to be collected is justified.
- There is an ongoing review process that considers whether a particular programme is still operating in the most cost-effective way to meet its stated requirements.
- Data collection instruments are designed to minimize coding and editing cost and time

Requirement 11.2: Procedures exist to assess and justify demands for new statistics against their cost.

- Demands for new statistics are regularly registered and assessed by statistical experts with regard to the proposed methodology and associated costs, and are discussed by management, based on inputs from users and in cooperation with other stakeholders.
- Before contemplating a new data collection, there are mechanisms to review whether already available data sources can be utilized with minimal impact on their purpose and quality.
- When introducing new statistics, a cost-benefit analysis is conducted.

Requirement 11.5: Proactive efforts are made to improve the statistical potential of administrative data and other data sources.

- Statistical agencies provide input to the legislative process to obtain and maintain access to administrative and other data sources for statistical purposes, if needed.
- Appropriate arrangements (e.g., service-level agreements or national legislation) with owners or holders of administrative data and other data collections are made and updated as needed, specifying the access to and flow of data and metadata and other relevant aspects.
- An assessment of possible administrative data sources is carried out prior to launching any new survey.
- Data linking and integration methods are proactively pursued while ensuring data security and privacy.
- Quality reports for administrative and other data used for official statistics are established by the responsible statistical agency in cooperation with the data owners or holders.

Requirement 12.1: Statistical processes are tested before implementation.

- The testing strategy is developed as part of the design phase of the statistical business process model.
- Data capture procedures and data collection tools and instruments such as electronic systems are tested to ensure simplicity and minimal intrusion on privacy, and are adjusted if required before their implementation.
- Survey questionnaires are tested using appropriate methods (e.g., pilot survey, focus groups, etc.).
- Collection systems for administrative and other data are tested before use.
- Data treatment and data processing procedures are tested and adjusted, if required and possible, prior to their actual application.
- Test results are taken into account in the implementation of the production process and are approved.
- In the case of integrating data from one or more sources, the quality of the linkage procedures is tested.

Requirement 12.2: Statistical processes are well established and regularly monitored and revised as required.

- The statistical agencies have documented procedures and guidelines that contain recommendations for appropriate methodologies to be used at different steps of the statistical production process.
- Documentation of production processes should follow the GSBPM.
- A policy for archiving data and statistics is in place and is followed.
- Statistical procedures employ internationally recognized statistical techniques.
- Data of all data sources are reviewed and validated to identify potential problems, errors and discrepancies such as outliers, missing data and miscoding.
- When coding is done through an automated process, a team of well-trained coders is assigned to verify the automated coding and to handle un-coded cases.
- The effects of data editing and imputation are analyzed as part of assessing the quality of the data collection.
- All statistical databases are designed and arranged in a way that allows and facilitates data linkage, using unique identifiers for statistical units as appropriate while ensuring data security and privacy.

Requirement 12.3: Procedures are in place to effectively use administrative and other data sources for statistical purposes.

- Statistical agencies use tools and guidelines to assess the quality of the data of administrative and other data sources.
- Appropriate processes and software applications for the collection, processing and analyses of data of administrative and other data sources have been developed and implemented.
- Owners or holders of administrative and other data sources inform the statistical agencies of any changes in the data production process.
- Metadata related to administrative or other data sources are available to the statistical agencies, including concepts and definitions, classifications, coverage compared to target population and other quality aspects.
- Documentation exists that describes how data from administrative and other sources meet the statistical requirements in terms of definitions, concepts and coverage, among other things.

Requirement 12.5: Metadata and documentation of methods and different statistical processes are managed throughout the processes and shared as appropriate.

- There is a policy on metadata documentation linked to the statistical production processes.
- The policies and standards for maintaining and updating metadata are followed.
- Work on preparing statistics and their related metadata should be done in parallel.
- Metadata are captured throughout the statistical business process following the GSBPM and stored in a metadata management system.
- Statistical methods and processes are documented in such a way that allows for the recreation of the entire statistical production process.

Requirement 13.4: Data sharing, data linkage and the use of administrative and other data sources are promoted to minimize respondent burden.

- Documentation of data already available within the NSS, including archived data, exists and is shared.
- Procedures and technical tools for data sharing and data linkage within the NSS (e.g., formal agreements, web services, common databases) exist.
- Data repositories are shared among statistical agencies for the production of official statistics and in compliance with confidentiality policies.
- Information on the quality of data to be linked exists (e.g., on coverage and linkage possibilities).

- Use of administrative and other data as an alternative to survey data for producing official statistics is promoted throughout the NSS

Requirement 14.1: Procedures are in place to identify users and their needs and to consult them about the content of the statistical work programme.

- There is legislation or some other formal provision which includes an obligation to consult with the main users of the statistics.
- Structured and periodic consultation processes (e.g., advisory councils and committees or working groups) with key stakeholders and users are in place to review the content of the statistical programme and the usefulness of existing statistics, and to identify requirements for new statistics.
- Feedback from a user support service, centre or hotline is analyzed to understand and identify user needs.
- Data on the use of statistics (e.g., web analytics, number and types of downloads, subscribers to reports) are collected and analyzed to improve statistical outputs.

Requirement 14.3: Statistics based on new and existing data sources are being developed in response to society's emerging information needs.

- An innovation laboratory is established to consider and experiment with new data sources to meet emerging information needs.
- Cooperation with the scientific community and owners or holders of new data sources is established to experiment with and pioneer the use of these data sources.
- Possibilities of exploiting new data sources are regularly discussed by management.

Requirement 14.4: User satisfaction is regularly measured and systematically followed up.

- User satisfaction surveys and user studies are regularly carried out and analysed.
- Improvement actions arising from the user satisfaction surveys and user studies are identified and implemented.
- User satisfaction surveys include questions on the opinions of users about metadata availability.
- Measures to assess the satisfaction of main users with particular products are in place (e.g., specific user-satisfaction surveys and indicators, including timeliness, etc., at the product level).

Requirement 15.1: Source data, integrated data, intermediate results and statistical outputs are regularly assessed and validated.

- Systems for assessing and validating source data, integrated data, intermediate results and statistical outputs are developed and managed.
- Data are systematically checked and compared with data from other sources and over time.
- Results of statistics are compared with other existing information in order to ensure validity.

Requirement 15.3: Studies and analyses of revisions are carried out and used to improve data sources, statistical processes and outputs.

- Preliminary and revised data and statistics are clearly identified.
- Explanations about the timing, reasons for and the nature of revisions are made available.
- The revision policy follows standard and transparent procedures.
- Information on the size and direction of revisions for key indicators is used to improve the statistical processes.
- Information on the size and direction of revisions for key indicators is provided and made public.

Requirement 16.2: The relationship with data providers is managed with regard to timeliness and punctuality needs.

- Agreements are in place with data providers on the planned delivery dates and delivery format.

- Procedures are in place to ensure the effective and timely flow of data from providers to statistical agencies.
- Follow-up procedures are in place to ensure the timely receipt of data from providers.

Requirement 17.4: Access to microdata is allowed for research purposes, subject to specific rules and protocols on statistical confidentiality that are posted on the statistical agency's website.

- The statistical agency controls or monitors the access of researchers to microdata by providing the microdata in a secure environment.
- Researchers are regularly consulted about the effectiveness of the microdata access arrangements.
- Remote access facilities are available for accessing microdata, with appropriate controls.

Requirement 18.2: Procedures or guidelines are in place to ensure and monitor internal, intrasectoral and cross-sectoral coherence and consistency.

- Statistics derived from different sources or with different periodicities (e.g., monthly, quarterly, yearly) are compared and any differences are explained and reconciled, as appropriate.
- Cooperation and the exchange of knowledge among individual statistical programmes and domains is promoted.
- Process-specific procedures and guidelines are available to ensure that outputs are internally coherent.
- Before new statistics or statistical programmes are launched, the conceptual and methodological relationship with existing statistics is analysed.
- Statistical outputs are compared with results of other statistical or administrative sources that provide the same or similar information on the same subject matter, and divergences are identified and explained to users.
- Internal procedures or guidelines are developed in order to ensure and monitor internal coherence and consistency.
- Procedures and guidelines are developed in order to ensure that results from different sources can be combined. Compliance is periodically assessed.

Requirement 19.2: Metadata are documented, archived and disseminated according to internationally accepted standards.

- International, regional, national or internal standards are used for metadata documentation, management and archiving.
- Procedures are in place to ensure that metadata are documented according to standardized metadata systems, and are regularly updated.
- Metadata are made available at the same time as the data and statistics to which they pertain.
- The dissemination of metadata is tailored to different needs, such as those of producers and users of statistics.
- A systematic way to archive metadata is available that also ensures that the metadata are accessible for reuse in the future.
- A glossary of statistical concepts is publicly available.

Annex 4: Mapping of the ten critical requirements (CR) for the quality assurance when using administrative and other data sources to the conceptual approach and relevant UN NQAF requirements

Table 3 shows the linkages between the ten critical requirements for quality assurance (QA) when using administrative and other data sources (AOS) and relevant UN NQAF requirements, following the conceptual approach. The table shows what is already covered in UN NQAF and hereby helps to avoid double work. It also shows gaps in UN NQAF with respect to QA for administrative and other data sources and can therefore be used to update UN NQAF and respective national frameworks. In addition, the table shows how the ten critical requirements relate to the actions and overarching considerations of the conceptual approach.

Table 3: Mapping of the ten critical requirements (CR) for quality assurance when using administrative and other data sources and relevant UN NQAF requirements, following the conceptual approach

Critical requirement for QA of AOS	Relevant UN NQAF requirement
Important preconditions: Data access and confidentiality	
CR 1: The use of administrative and other data sources must meet a set of preconditions and take user needs into consideration	Requirements 2.5, 2.6 Requirements 7.1 – 7.6 Requirements 14.1, 14.3
Action 1: Identification of statistical need and selection of data source	
CR 2: New data sources, data providers as well as the use of multiple data sources are proactively explored	Requirements 10.3, 10.5 Requirement 11.1, 11.2, 11.5 Requirements 14.1, 14.3 Requirement 5.3
CR 3: There is basic information about the data provider and general information about the data source	Requirement 10.3
CR 4: The data provider and data source are assessed for their risks	Requirement 10.3 Requirement 5.2, 5.3 Requirement 7.5
CR 7: The quality of the input data is systematically evaluated by the statistical agency	Requirement 10.3 Requirement 12.3 Requirement 15.1
CR 8: There is comprehensive metadata about the input data	Requirement 10.3 Requirement 12.3, 12.5
Action 2: Cooperation with data provider	
CR 4: The data provider and data source are assessed for their risks	Requirement 10.3 Requirement 7.5
CR 5: There are cooperation agreements with the data providers and there is ongoing cooperation, as applicable	Requirements 2.5, 2.6, 2.7 Requirement 3.2 Requirement 16.2
CR 6: The data provider assures the quality of its data and produces a quality report (or quality declaration) in cooperation with the statistical agency, as applicable	Requirement 2.7 Requirement 10.3 Requirement 11.5
Action 3: Data acquisition and processing	
CR 7: The quality of the input data is systematically evaluated by the statistical agency	Requirement 10.3 Requirement 12.3

Critical requirement for QA of AOS	Relevant UN NQAF requirement
	Requirement 15.1
CR 8: There is comprehensive metadata about the input data	Requirement 10.3 Requirement 12.3, 12.5
CR 9: Processing of input data at the statistical agency follows standards, guidelines, and best practices	Requirement 12.1, 12.2, 12.3 Requirement 8.5 Requirement 10.1, 10.5 Requirement 11.5 Requirement 13.4 Requirement 15.3 Requirement 18.2
Action 4: Dissemination	
CR 8: There is comprehensive metadata about the input data	Requirement 12.3, 12.5
CR10: The dissemination of statistical outputs meets the needs of users	Requirements 14.1, 14.3 Requirements 15.1 Requirements 17.4 Requirements 19.2
Overarching consideration 1: User needs for relevant statistics*	
CR 1: The use of administrative and other data sources must meet a set of preconditions and take user needs into consideration	Requirements 14.1, 14.3
CR10: The dissemination of statistical outputs meets the needs of users	Requirements 14.1, 14.3 Requirements 17.4 Requirements 19.2
Overarching consideration 2: Quality of input data and metadata*	
CR 6: The data provider assures the quality of its data and produces a quality report (or quality declaration) in cooperation with the statistical agency, as applicable	Requirement 2.7
CR 7: The quality of the input data is systematically evaluated by the statistical agency	Requirement 10.3 Requirement 12.3 Requirement 15.1
CR 8: There is comprehensive metadata about the input data	Requirement 10.3 Requirement 12.3, 12.5
CR 9: 9. Processing of input data at the statistical agency follows standards, guidelines, and best practices	Requirement 12.1, 12.2, 12.3 Requirement 8.5 Requirement 10.1, 10.5 Requirement 11.5 Requirement 13.4 Requirement 15.3 Requirement 18.2
CR10: The dissemination of statistical outputs meets the needs of users	Requirements 14.1, 14.3 Requirements 17.4 Requirements 19.2

* Not all UN NQAF requirements that can be mapped to a specific critical requirement apply to the overarching considerations.
